Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPSP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PMO-HT0302-
081199-002-g02&t3=1999-11-08&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 41
High quality sequence stop: 590.

FEATURES   Location/Qualifiers
     source   1..608
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0302"
                     /note="Organ: head_neck; Vector: puc18; Site 1: SmaI;
                     Site 2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No.196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."

ORIGIN

Query Match   18.2%; Score 453; DB 10; Length 608;
Best Local Similarity 98.9%; Pred. No. 3.8e-77;
Matches 456; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy   835 TACCTTACAGTTAATCCATGGAATCAATCAAATTAAATCATCATGACTACATTGATTC 894
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   605 TAGATACCAGTTAATCCATGGAATCAATCAAATTAAATCATCATGACTACATTGATTC 546

Qy   895 CCCATCGCTGAAGGACAGAATTCATTGTGTGGCATTGTATTGTGATGCCAGCTCTATTCA 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   545 CCCATCGCTGAAGGACAGAATTCATTGTGTGGCATTGTATTGTGATGCCAGCTCTATTCA 486

Qy   955 ATACTTCTCCTCTCAGATGATAGTAAAGATCAAAGAATTCGAAGGAGTTGGTAAACGC 1014
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   485 ATACTTCTCCTCTCAGATGATAGTAAAGATCAAAGAATTCGAAGGAGTTGGTAAACGC 426

Qy   1015 TGGTGTGGTACATGTGGCTTGCTCACTCATGTGGATAGCATGGATTTGATTACAAAAGG 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   425 TGGTGTGGTACATGTGGCTTGCTCACTCATGTGGATAGCATGGATTTGATTACAAAAGG 366

Qy   1075 TGACCTTATAGAAATAGAAGATGTGAGCCTGAGCCTGAGGCTCCAAGCTAGAGGAAGTCCAAAG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   365 TGACCTTATAGAAATAGAAGATGTGAGCCTGAGCCTGAGGCTCCAAGCTAGAGGAAGTCCAAAG 306

Qy   1135 AAAACTTGGATTTGCTCTTTCTGACATCCGGTGGTTAGCAATTATTCCTCCGAGTGGGA 1194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   305 AAAACTTGGATTTGCTCTTTCTGACATCCGGTGGTTAGCAATTATTCCTCCGAGTGGGA 246

Qy   1195 GCTGGACCCTGTAAAGGATGTTCAATTCTTTCTGAGACGAATGCTATGGGCTGC 1254
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   245 GCTGGACCCTGTAAAGGATGTTCAATTCTTTCTGAGACGAATGCTATGGGCTGC 186

Qy   1255 AGATGACTTCTTAGAGGATTTGCCTTTTGAGCAAATAGGTA 1295
          |||||||||||||||||||||||||||||||||||||||||
Db   185 AGATGACTTCTTAGAGGATTTGCCTTTTGAGCAAATAGGGA 145

RESULT 9
AI275105/c
LOCUS   AI275105   451 bp   mRNA   linear   EST 29-JAN-1999
DEFINITION ql65d11.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1877205
            3', mRNA sequence.
ACCESSION   AI275105

                     /clone="IMAGE:470583"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Soares pregnant uterus NbHPU"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I;
                     Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by M. Fatima Bonaldo."

ORIGIN

Query Match   19.0%; Score 472.8; DB 9; Length 504;
Best Local Similarity 98.6%; Pred. No. 6e-81;
Matches 496; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

Qy   1977 CAAATTGCTCGGGAAACTCTCATGTTA-CTAACTTGTTGTGCTCTAACTCTGCCATCTTGGT 2035
          ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   503 CAAATTGCTCGGGAAACTCTCANGTTACCAACTTTGTTGCTCTAACTCTGCCATCTTGGT 444

Qy   2036 TTCCCCATCCCTTCTCTTCCTCATGGTACGTGCGTGCTCCTAATATATAGGCTTGGTTGGAGAT 2095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   443 TTCCCCATCCCTTCTCTTCCTCATGGTACGTGCTCCTAATATATTNGGCGTTGGTTGGAGAT 384

Qy   2096 TTTCAGTGGGTCC-AATATTCCTCTCTCCCTCTGGTTGCCTTTCCTGAGATAATCCACTAAG 2154
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   383 TTTCAGTGGTCNAATATACCCTCTTTCCCTCTGGTTGCCTTTCCTGAGATAATCCACTAAG 324

Qy   2155 AATATTTTGTGTTTCTTTTCTTCCAGGGAATCTCAAGGAGGAAATTATCAACTGTGCACAAG 2214
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   323 AATATTTTGTGTTTCTTTTCTTCCAGGGAATCTCAAGGAGGAAATTATCAACTGTGCACAAG 264

Qy   2215 GAAAAAAATAGATATGTGAAAGGTTCACGTAAATTCCTCCACATCACAGAAGATTAAAAT 2274
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   263 GANAAAAATAGATATGTGAAAGGTTCACGTAAATTTCCTCCACATCACAGAAGATTAAAAT 204

Qy   2275 TCAGAAAGGAGAAAAACACAGACCAAAGAAGTATCTAAGACCAAAGGGATGTGTTTTAT 2334
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   203 TCAGAAAGGAGAAAAACACAGACCAAAGAAGTATCTAAGACCAAAGGGATGTGTTTTAT 144

Qy   2335 TAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTACTTGTAAATAACTAGAAATAA 2394
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   143 TAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTACTTGTAAATAACTAGAAATAA 84

Qy   2395 CATGATTAGTCATAATTGTGAAAAATAATAATTTTCTTGGATTTATTGTTCTGTAT 2454
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   83 CATGATTAGTCATAATTGTGAAAAATAATAATTTTTCTTGGATTTATTGTTCTGTAT 24

Qy   2455 CTGTGAAAAAAAAATAAAATTTCTTAT 2477
          ||||||||||||||||||||||||||
Db   23 CTGTGAAAAAAAAAATAAAATTTCTTAT 1

RESULT 8
AW381442/c
LOCUS   AW381442   608 bp   mRNA   linear   EST 04-FEB-2000
DEFINITION PMO-HT0302-081199-002-g02 HT0302 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW381442
VERSION   AW381442.1 GI:6886101
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 608)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE   The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Simpson A.J.G.

VERSION   AI275105.1 GI:3897379
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 451)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Insert Length: 932   Std Error: 0.00
             Seq primer: -40UP from Gibco.
FEATURES   Location/Qualifiers
     source   1..451
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1877205"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /clone_lib="Soares NhHMPu S1"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site 1: Not I;
                     Site 2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
ORIGIN

Query Match   18.1%;   Score 450; DB 9; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.5e-76;
Matches   450; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   2035 TTTCCCATCCCTTCTCTTCCTTCCTTCCCTCTGGTTGCCTTTCCTGAGATAATCCACTAAG 2094
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   451 TTTCCCATCCCTTCTCTTCCTTCCTTCCCTCTGGTTGCCTTTCCTGAGATAATCCACTAAG   392

Qy   2095 TTTTCAGTGGTCCAATATTCCTTCCCTTCCCTTGCCTTTCCTGAGATAATCCACTAAG 2154
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   391 TTTTCAGTGGTCCAATATTCCTTCCCTTCCCTTGCCTTTCCTGAGATAATCCACTAAG   332

Qy   2155 AATATTTTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGAAATTATCAACTGTGCACAAG 2214
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   331 AATATTTTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGAAATTATCAACTGTGCACAAG   272

Qy   2215 GAAAAAATAGATATGTGAAAGGTTCACGTAAATTTCCTCACATCACAGAAGATTAAAAT 2274
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   271 GAAAAAATAGATATGTGAAAGGTTCACGTAAATTTCCTCACATCACAGAAGATTAAAAT   212

Qy   2275 TCAGAAAGGAGAAAACACAGACCAAAGAGTATCTAAGACCAAAGGGATGTGTTTTAT 2334
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211 TCAGAAAGGAGAAAACACAGACCAAAGAGTATCTAAGACCAAAGGGATGTGTTTTAT   152

Qy   2335 TAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTACTTGTAAATAACTAGAAATAA 2394
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   151 TAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTACTTGTAAATAACTAGAAATAA   92

Qy   2395 CATGATTTAGTCATAATGTGAAAAATAATAAATTTTCTTGGATTTATGTTCGTAT 2454
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91 CATGATTTAGTCATAATGTGAAAAATAATAAATTTTCTTGGATTTATGTTCGTAT   32

Qy   2455 CTGTGAAAAATAAATTCTTATAAAACTC 2484

Db   31 CTGTGAAAAATAAATTCTTATAAAACTC 2
           ||||||||||||||||||||||||||||

RESULT 10
AL541045/c
LOCUS   AL541045   1201 bp   mRNA   linear   EST 12-MAY-2003
DEFINITION AL541045 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE005YC19
            3-PRIME, mRNA sequence.
ACCESSION   AL541045
VERSION   AL541045.2 GI:30544837
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On Feb 15, 2001 this sequence version replaced gi:12871741.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 1009.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DE005AB10NP1&cluster=1009.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DE005AB10NP1.
FEATURES   Location/Qualifiers
     source   1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE005YC19"
                     /tissue_type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

Query Match   17.8%;   Score 442.6; DB 9; Length 1201;
Best Local Similarity 90.5%; Pred. No. 3.4e-75;
Matches   448; Conservative   19; Mismatches   28; Indels   0; Gaps   0;

Qy   801 AAATATTTTTGAATTAATATCTTGTCTTTATGTCTCACCTTACAGTTAATCCATGGAAT   860
           ||||| ||| | |||| ||||||| |||| |||||||||||||||||||||||||||
Db   791 ACATATTCTATATCTTGAACGTAACATTCCTGATAGATACCAGTTTAATCCCATGGAAT   732

Qy   861 CAATCAAATAAATCATCATGACTACATTGATTCCCATGCTGAAGGACAGAATTCATT   920
           |||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   731 CAATCAAATAAATCATCATGACTACATTGATTCCCATGCTGAAGGACAGAATTCATT   672

Qy   921 GTGTGGCATTGTATTGATGCGCCAGCTCATTCAATCAATCAATCTCAGATGAGTAGTAA   980
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   671 GTGTGGCATTGTATTGATGCGCCAGCTCATTCAATCAATCAATCTCAGATGAGTAGTAA   612

Qy   981 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGGTGGTGGCTTTGCTCTCA 1040
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   611 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGGTGGTGGCTTTGCTCTCA   552

Qy   1041 CTCATGTGGATAGCATGGATTTGATTACAAAAAGGTGACCTTATAGAAAAATAGAGATGTG 1100
           |||||||||||||||||||||||||||||||||||| | |||||||||||||||||||
Db   551 CTCATGTGGATCGCATGGATTTGATTGATTYMCAAAAGTGCCCTTATMGAAAATMGAGATGTG   492

Qy   1101 AGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACA 1160

RESULT 11

```
Qy   491  CGCCTGTGMGGTCCAMGCTMGMGGAAGTCCAAMGAAMACTTGGATTTGCTCTTTCTGACM  432
          |||||||:|||||:|||:||:|:|||| |:||| ||||||||:|||:|| ||||||||||
Db

Qy  1161  TCTCCGGTGGTTAGCAATTATTCCTCTGAGTGGGGAGCTGGACCCTGTAAAGGATGTTCTAA  1220
          |||||||||||:||||||:|:||||||||||:|||||||||||||:|| |||:|||||||||
Db   431  TCTCCGGTGGTTCGCAATTATTCCTCTGAGTGGMGCTGGGMGCCCTGTAAMGGATGTTCTAA   372

Qy  1221  TTCTTTCTGCTCTGAGACGACGAATGCTATGGGCTGCAGATGACTTCTTAGGAGATTTGCCTT  1280
          |||||||||||||||||||||:|||| ||||||||||||||||||||||:||||||||||||
Db   371  TTCTTTCTGCTCTGAGACGACGAATGCTATGGGCTGCMGATGACTTCTTMGMGGAGATTTGCCTT   312

Qy  1281  TTGAGCAAATAGGTA  1295
          |||:||:|||||||
Db   311  TTGMGCAAATAGGGA  297
```

```
RESULT 11
AW130518/c
LOCUS       AW130518    xf52d08.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2621679 3',
            452 bp  mRNA   linear   EST 27-OCT-1999
DEFINITION  mRNA sequence.
ACCESSION   AW130518
VERSION     AW130518.1  GI:6132125
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 452)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1..452
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2621679"
                     /tissue_type="poorly differentiated adenocarcinoma with
                     signet ring cell features"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Gas4"
                     /note="Organ: stomach; Vector: pCMV-SPORT6; Site 1: SalI;
                     Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.69 kb. Life Technologies catalog #:
                     11549-011"
ORIGIN
```

```
Query Match       17.7%;  Score 440.8;  DB 10;  Length 452;
Best Local Similarity  99.5%;  Pred. No. 9e-75;
Matches  442;  Conservative  0;  Mismatches  2;  Indels  2;  Gaps  0;

Qy  2041  CATCCCTTCCTTCCTTCCTCATGGTAGGTAGCTGTGTCTCCTAATATTAGCGTTGGTTGAGATTTTCA  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   452  CATCCCTTCCTTCCTTCCTCATGGTAGCGTGTCCGTGTCCTAATATTAGCGTTGGTTGAGATTTTCA   393

Qy  2101  GTGGTCCAATATTCCTCTTCCCCTCTGGTTGCCTTTCCTGAGATAATCCACTAAGAATATT  2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   392  GTGGTCCAATATTCCTCTTCCCCTCTGGTTGCCTTTCCTGAGATAATCCACTAAGAATATT   333
```

RESULT 12

```
Qy  2161  TTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGAGGAAATTATCAACTGTGCACAAGGAAAAA  2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   332  TTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGAGGAAATTATCAACTGTGCACAAGGAAAAA   273

Qy  2221  AATAGATATGTGAAAGGTTCACGTAAATTCCTCACATCACAGAAGATTAAAATTCAGAA  2280
          |||||||||||||||||||||:||||||||||||||:|||||||||||||||||||||
Db   272  AATAGATATGTGAAAGGTTCACGTAAATTCCTCACATCACAGAAGATTAAAATTCAGAA   213

Qy  2281  AGGAGAAAACACAGACCAAAGGATGTGTTTTATTAATGT  2340
          |||||||||||||||||||:|||||||||||||||||||
Db   212  AGAAGAAAACACAGACCAAAGAAGGGATGTGTTTTATTAATGT   153

Qy  2341  CTAGGATGAAGAAATGCATAGAACATGTAGTAGTACTTGTAAATAACATGAT  2400
          ||||||||||||||||||||||||||||||||||||||:||||||||||||
Db   152  CTAGGATGAAGAAATGCATAGAACATTGTAGTACTTGTAAATAACATGAT   93

Qy  2401  TTAGTCATAATTGTGAAAAATAATAATAATTTTCTTGGATTTATGTTCTGTATCTGTGA  2460
          ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db    92  TTAGTCATAATTGTGAAAAATAATAATAATTTTCTTGGATTTATGTTCTGTATCTGTGA   33

Qy  2461  AAAAATAAATTCTTATAAAACTC  2484
          ||||||||||||||||||||||
Db    32  AAAAATAAATTCTTATAAAACTC   9
```

```
RESULT 12
AI290106/c
LOCUS       AI290106    qw26d02.x1 NCI_CGAP_Ut4 Homo sapiens cDNA clone IMAGE:1992195 3',
            436 bp  mRNA   linear   EST 21-DEC-1998
DEFINITION  mRNA sequence.
ACCESSION   AI290106
VERSION     AI290106.1  GI:3931772
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 436)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 765    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 405.
FEATURES             Location/Qualifiers
     source          1..436
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1992195"
                     /tissue_type="serous papillary carcinoma, high grade, 2
                     pooled tumors"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Ut4"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;
                     Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.48 kb. Life Technologies catalog #:
                     11542-016"
ORIGIN
```

```
Query Match       17.6%;  Score 436;  DB 9;  Length 436;
Best Local Similarity  100.0%;  Pred. No. 7.6e-74;
```

```
Matches 436;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  2049  CTCTTCCTCATGGTACGTGCGTCCTAATATTAGCGTTGGTTGGATTTTCAGTGGTCCA  2108
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   436  CTCTTCCTCATGGTACGTGCGTCCTAATATTAGCGTTGGTTGGATTTTCAGTGGTCCA   377

Qy  2109  ATATTCCTCTCCCTCGGTTGCCTTCTGAGATAATCCACTAAGAGATATTTGTGTTT  2168
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   376  ATATTCCTCTCCCTCGGTTGCCTTCTGAGATAATCCACTAAGAGATATTTGTGTTT   317

Qy  2169  CTTTTCTCAGGGAATCTAAGGAGGAAATTATCAACTGTGCACAAGGAAAAAATAGATA  2228
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   316  CTTTTCTCAGGGAATCTAAGGAGGAAATTATCAACTGTGCACAAGGAAAAAATAGATA   257

Qy  2229  TGTGAAAGTTCACGTAAATTTCCTCACATCACAGAGATTAAAATTCAGAAAGGAGAAA  2288
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   256  TGTGAAAGTTCACGTAAATTTCCTCACATCACAGAGATTAAAATTCAGAAAGGAGAAA   197

Qy  2289  ACACAGACCAAGAGAAGTATCTAAGACAAGGGGATGTGTTTTATTAATGTCTAGGATG  2348
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   196  ACACAGACCAAGAGAAGTATCTAAGACAAGGGGATGTGTTTTATTAATGTCTAGGATG   137

Qy  2349  AAGAAATGCATAGAACATTGTAGTACTTGTAAATAACATGAGAAATAACATGATTTAGTCAT  2408
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   136  AAGAAATGCATAGAACATTGTAGTACTTGTAAATAACATGAGAAATAACATGATTTAGTCAT   77

Qy  2409  AATTGTGAAAAATAATAATAATTTTTCTTGGATTTATGTTCTGTATCTGTGAAAAAATAA  2468
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    76  AATTGTGAAAAATAATAATAATTTTTCTTGGATTTATGTTCTGTATCTGTGAAAAAATAA   17

Qy  2469  ATTTCTTATAAAACTC  2484
          ||||||||||||||||
Db    16  ATTTCTTATAAAACTC     1
```

RESULT 13
N33172/c

LOCUS	N33172 462 bp mRNA linear EST 10-JAN-1996
DEFINITION	yy35c11.s1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:273236 3', mRNA sequence.
ACCESSION	N33172
VERSION	N33172.1 GI:1153571
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 462)
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE	The WashU-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 365 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: m13 -40 forward High quality sequence stop: 365.
FEATURES	Location/Qualifiers
source	1..462 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="GDB:3882878" /db_xref="taxon:9606"

```
                          /clone="IMAGE:273236"
                          /sex="Male"
                          /tissue_type="melanocyte"
                          /lab_host="DH10B (ampicillin resistant)"
                          /clone_lib="Soares melanocyte 2NbHM"
                          /note="Vector: pT7T3D (Pharmacia) with a modified
                          polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                          was primed with a Not I - oligo(dT) primer [5'
                          TGTTACCAATCTGAAGTGGGAGCGGCCGCCAGTTTTTTTTTTTTTTTTTT 3'],
                          double-stranded cDNA was size selected, ligated to Eco RI
                          adapters (Pharmacia), digested with Not I and cloned into
                          the Not I and Eco RI sites of a modified pT7T3 vector
                          (Pharmacia). Library constructed by Bento Soares and
                          M.Fatima Bonaldo. RNA from normal foreskin melanocytes
                          (FS374) was kindly provided by Dr. Anthony P. Albino."
ORIGIN
```

```
Query Match             17.5%;  Score 434.8;  DB 14;  Length 462;
Best Local Similarity   99.1%;  Pred. No. 1.3e-73;
Matches  458;  Conservative   0;  Mismatches  2;  Indels  2;  Gaps  2;

Qy  2018  TAACTCTGCCATCTTGGTTCCCCAT-CCCTTCCTCCTCCATGGTACGTGCGTCCTAA  2076
          |||||||||||||||||||||||| | ||||||||||||||||||||||||||||||
Db   462  TAACTCTGCCATCTGGGTTCCCCATCCCCATCCCCTTCCTCCTCCATGGTACGTGCGTCCTAA   403

Qy  2077  TATTAGCGTTGGTTGGAGATTT-CAGTGGTCCAATATTCCTCTCCCTCGGTTGCCTTT  2135
          ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db   402  TATTAGCGTTGGTTGGAGATTTTTCCAGTGGTCCAATATTCCTCTCCCTCGGTTGCCTTT   343

Qy  2136  CCTGAGATAATCCACTAAGAATATTTGTGTTTCTTTTCTCAGGGAATCTAAGGAGGAA  2195
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   342  CCTGAGATAATCCACTAAGAATATTTGTGTTTCTTTTCTCAGGGAATCTAAGGAGGAA   283

Qy  2196  ATTATCAACTGTGCACAAGGAAAAAATAGATATGTGAAAGTTCACGTAAATTCCTCA  2255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   282  ATTATCAACTGTGCACAAGGAAAAAATAGATATGTGAAAGTTCACGTAAATTCCTCA   223

Qy  2256  CATCACAGAAGATTAAAATTCAGAAAGGAGAAAAACACAGACCAAGAGAAGTATCTAAGA  2315
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   222  CATCACAGAAGATTAAAATTCAGAAAGGAGAAAAACACAGACCAAGAGAAGTATCTAAGA   163

Qy  2316  CCAAAGGGATGGTGTTTTTATTAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTACT  2375
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   162  CCAAAGGGATGGTGTTTTTATTAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTACT   103

Qy  2376  TGTAAATAACTAGAAATAACATGATTTAGTCATAATTGTGAAAAATAATAATTTTTC  2435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102  TGTAAATAACTAGAAATAACATGATTTAGTCATAATTGTGAAAAATAATAATTTTTC    43

Qy  2436  TTGGATTTATGTTCTGTATCTGTGAAAAATAAATTTCTTAT  2477
          |||||||||||||||||||||||||||||||||||||||||
Db    42  TTGGATTTATGTTCTGTATCTGTGAAAAATAAATTTCTTAT     1
```

RESULT 14
CA411843/c

LOCUS	CA411843 763 bp mRNA linear EST 07-NOV-2002
DEFINITION	UI-H-EZ0-bak-o-13-0-UI.s1 NCI_CGAP_Ch1 Homo sapiens cDNA clone UI-H-EZ0-bak-o-13-0-UI 3', mRNA sequence.
ACCESSION	CA411843
VERSION	CA411843.1 GI:24774494
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 763)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov

RESULT 15

```
LOCUS       AA032029               512 bp    mRNA    linear   EST 09-MAY-1997
DEFINITION  zk15b08.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:470583 5', mRNA sequence.
ACCESSION   AA032029
VERSION     AA032029.1  GI:1501992
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
  PUBMED    8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 670   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 351.
FEATURES             Location/Qualifiers
     source          1..512
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3756585"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:470583"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5,
                     AACTGGAAGAATTCGCGGCCGCGCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by M. Fatima Bonaldo."
ORIGIN
```

```
Query Match             17.4%;    Score 431;  DB 9;  Length 512;
Best Local Similarity   97.2%;    Pred. No. 6.8e-73;
Matches   489;  Conservative    0;  Mismatches    9;  Indels    5;  Gaps    5;

Qy     1923 TAGTGCTTGGGAAATCAATTCAGAATATCACATCCTTCTCTTACTCTCCAATTCTCTCTACTCTCCAAATT 1982
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 TAGTGCTTGGGAAATCAATTCAGAATATCACATCCTTCTCTCTACTCTCCAATTCTCTCTACTCTCCAAATT 60

Qy     1983 GCTGGGAAACTCTCATGTTACTAACTTTGTTGCTCCATCCTGGTTCCCCA 2042
            |||||||||||||||||||||||||||||||||||||||||||
Db       61 GCTGGGAAACTCTCATGTTACTAACTTCTTGTTGCTCCATCTGGTTGGTTCCCCA 120

Qy     2043 TCCCTTCTTCTCCTCCATGGTACGTGTGCTCCTAATATTAGCGGTGGTTGAGATTTTCAGT 2102
            |||||||||||||||||||||||||||||||||||||||||||||||||
```

Tissue Procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of
Orthoapedics
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=Yes.

```
FEATURES             Location/Qualifiers
     source          1..763
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-EZ0-bak-o-13-0-UI"
                     /tissue_type="Chondrosarcoma Grade II"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_Ch1"
                     /note="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia)
                     with a modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_Ch1 is a cDNA library containing the following
                     tissue(s): Chondrosarcoma Grade II. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     TGATCACGCT.
                     TAG_TISSUE=grade-2-chondrosarcoma
                     TAG_LIB=UI-H-EZ0
                     TAG_SEQ=ATCTTAATATG"
ORIGIN
```

```
Query Match             17.4%;    Score 433.4;  DB 14;  Length 763;
Best Local Similarity   99.5%;    Pred. No. 2.2e-73;
Matches   434;  Conservative    0;  Mismatches    0;  Indels    .2;  Gaps    0;

Qy      860 TCAATCAAATTAAATCATCATGACTACATCATTGATTCCCATCGCTGAAGGACAGAATTCAT 919
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      763 TCAATCAAATTAAATCATCATGACTACATCATTGATTCCCATCGCTGAAGGACAGAATTCAT 704

Qy      920 TGTGTGGCATTGTATTGATTGATGCCAGCTCATTCAATACTTCTCCTTCCAGATGATAGTA 979
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      703 TGTGTGGCATTTGTATNTGATGCCAGCTCATTCAATACTTCTCCTTCCAGATGATAGTA 644

Qy      980 AAGATCAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGGTACATGTGGCTTTGCTC 1039
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      643 AAGATCAAAGAATTCGAAGGAGTTGGTAAACGCTGGTGTGGTACATGTGGCTTTGCTC 584

Qy     1040 ACTCATGTGGATAGCATGGATTTGATTACAAAGGTGACCTTATAGAAAATAGAGAGATGT 1099
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db      583 ACTCATGTGGATAGCATGGATTTGATTACAAAGGTGACCTTATAGAAAATAGAGAGATGT 524

Qy     1100 GAGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGAC 1159
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db      523 GAGCCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGAC 464

Qy     1160 ATCTCGGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTA 1219
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db      463 ATCTCGGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTA 404

Qy     1220 ATTCTTTTCTGCTCTGAGACGAATGCATGGGCTGCAGATGACTTCTTAGAGGATTTGCCT 1279
            ||||||||||||||||||||||||||||||||||||||||||||||
Db      403 ATTCTTTTCTGCTCTGAGACGAATGCATGGGCTGCAGATGACTTCTTAGAGGATTTGCCT 344

Qy     1280 TTTGAGCAAATAGGTA 1295
            ||||||||||||
Db      343 TTTGAGCAAATAGGGA 328
```

```
Db   121  TCCCTTCTTCCTCCTCATGGTACGTGTGCTCCTAATATTAGCGTTGGTTGAGATTTTCAGT 180

Qy  2103  GGTCCAATATTCCTCCTTCCCCTCTGGGTTGCCTTTCCTGAGATAATCCACTAAGAATATTTT 2162
          |||||||||| ||| |||||| |||||||||||||||||||||||||||||||||||||||
Db   181  GGTCCAATATACCTCTTCCTCCTGGTTGCCTTTCCTGAGATAATCCACTAAGAATATTTT 240

Qy  2163  GTGTTTCTTTCTCAGGGAATCTAAGGAGGAAATTATCAACTGTGCACAAGGAAAAAAAA 2222
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GTGTTTCTTTCTCAGGGAATCTAAGGAGGAAATTATCAACTGTGCACAAGGAAAAAAA 300

Qy  2223  TAGATATGTGAAAGGTTCACGTAAATTTCCTCACATCACAGAAGATTAAAATTCAGAAAG 2282
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  TAGATATGTGAAAGGTTCACGTAAATTTCCTCACATCACAGAAGATTAAAATTCAGAAAG 360

Qy  2283  GAGAAAACACAGACCAA-AGAGAAGTATCTAAGACCAAAGGGATGTGTTTATTAAT-GT 2340
          |||||| |||||||||| ||||||||||||||||||||||||||||||| |||||| ||
Db   361  GAGAAANACAGACCAAGAGAGAAGTATCTAAGACCAAAGGGATGTGTTTTATTAATGGT 420

Qy  2341  CTAGGATGAAGAAATGCATA-GAACATTGTA-GTACTTGTAAATAACTAGAAATAACATG 2398
          |||||||||||| ||||||| |||||||||| ||||||||||||| |||||||||
Db   421  CTAGGATGAAGANATGCATAGGAACATTGTAGGTACTTGTAAATAACTAGAAATACCTGG 480

Qy  2399  ATTTA-GTCATAATTGTGAAAAA 2420
          ||||| ||||||||||||| ||||
Db   481  NTTTAGGTCATAATTGTGGAAAA 503
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    April 25, 2004, 07:15:18 ; Search time 123.228 Seconds
                                     (without alignments)
                                     11186.557 Million cell updates/sec

Title:          US-10-051-835-18
Perfect score:  2484
Sequence:       1 gtttattcctgtattataac........ataaattcttataaaactc 2484

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:   1365418

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :  Issued_Patents_NA:*
     1: /cgn2_6/ptcdata/2/ina/5A_COMB.seq:*
     2: /cgn2_6/ptcdata/2/ina/5B_COMB.seq:*
     3: /cgn2_6/ptcdata/2/ina/6A_COMB.seq:*
     4: /cgn2_6/ptcdata/2/ina/6B_COMB.seq:*
     5: /cgn2_6/ptcdata/2/ina/PCTUS_COMB.seq:*
     6: /cgn2_6/ptcdata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	240	9.7	240	4	US-09-397-787-62	Sequence 62, Appl
	2	196.6	7.9	2494	4	US-09-566-921-125	Sequence 125, Appl
	3	90.6	3.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	4	53.6	2.2	551	3	US-09-385-982-122	Sequence 122, App
c	5	48	1.9	6306	4	US-10-204-708-49	Sequence 49, Appl
	6	47.8	1.9	53332	4	US-09-801-861-3	Sequence 3, Appli
	7	45.8	1.8	1353	4	US-09-601-198-37	Sequence 37, Appl
c	8	45.2	1.8	660	1	US-07-991-867B-32	Sequence 32, Appl
c	9	45.2	1.8	660	1	US-08-107-755A-32	Sequence 32, Appl
	10	45.2	1.8	660	2	US-08-544-332-32	Sequence 32, Appl
	11	45.2	1.8	660	4	US-09-370-861A-32	Sequence 32, Appl
c	12	45.2	1.8	1511	1	US-07-991-867B-8	Sequence 8, Appli
c	13	45.2	1.8	1511	1	US-08-107-755A-8	Sequence 8, Appli
	14	45.2	1.8	1511	2	US-08-544-332-8	Sequence 8, Appli
c	15	45.2	1.8	1511	4	US-09-370-861A-8	Sequence 8, Appli
	16	45.2	1.8	4810	3	US-08-852-629-11	Sequence 11, Appl
	17	45.2	1.8	4838	3	US-08-852-629-15	Sequence 15, Appl
c	18	45.2	1.8	7218	1	US-08-232-463-14	Sequence 14, Appl
	19	44.2	1.8	8961	4	US-10-204-708-80	Sequence 80, Appl
	20	44.2	1.8	20674	4	US-09-641-638-651	Sequence 651, App
c	21	44	1.8	5219	4	US-10-204-708-52	Sequence 52, Appl
c	22	43.8	1.8	19513	4	US-10-204-708-39	Sequence 39, Appl
c	23	43.4	1.7	20674	4	US-09-641-638-651	Sequence 651, App
	24	43	1.7	832	4	US-09-621-976-2813	Sequence 2813, App
c	25	43	1.7	11049	4	US-10-204-708-22	Sequence 22, Appl
c	26	42.8	1.7	832	4	US-09-621-976-2813	Sequence 2813, Ap
c	27	42.8	1.7	5217	4	US-09-100-703A-25	Sequence 25, Appl
c	28	42.6	1.7	6020	4	US-10-204-708-8	Sequence 8, Appli
c	29	42.6	1.7	640681	4	US-09-790-988-1	Sequence 1, Appli
	30	42.4	1.7	658	3	US-08-998-416-595	Sequence 595, App
c	31	42.4	1.7	6156	4	US-10-204-708-60	Sequence 60, Appl
c	32	42	1.7	8920	2	US-08-446-855A-1	Sequence 1, Appli
c	33	42	1.7	8920	3	US-09-150-741-1	Sequence 1, Appli
	34	41.8	1.7	1667	4	US-08-485-284A-1	Sequence 1, Appli
c	35	41.8	1.7	148567	4	US-09-801-876B-3	Sequence 3, Appli
c	36	41.8	1.7	148567	4	US-10-254-869-3	Sequence 3, Appli
c	37	41.6	1.7	396	4	US-09-640-173-53	Sequence 53, Appl
c	38	41.6	1.7	396	4	US-09-713-550-53	Sequence 53, Appl
	39	41.6	1.7	6182	4	US-10-204-708-87	Sequence 87, Appl
	40	41.2	1.7	5915	4	US-10-204-708-77	Sequence 77, Appl
	41	41.2	1.7	10467	4	US-10-204-708-2	Sequence 2, Appli
c	42	41	1.7	11131	4	US-10-204-708-28	Sequence 28, Appl
c	43	40.8	1.6	782	3	US-08-998-416-224	Sequence 224, App
c	44	40.8	1.6	640681	4	US-09-790-988-1	Sequence 1, Appli
c	45	40.6	1.6	2539	3	US-08-749-522-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-09-397-787-62
; Sequence 62, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.466C2
; CURRENT APPLICATION NUMBER: US/09/397,787
; CURRENT FILING DATE: 1999-09-16
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 62
; LENGTH: 240
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-397-787-62

  Query Match             9.7%;  Score 240;  DB 4;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-49;
  Matches  240;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1393 GTTGTTGCCCTGTGATTAGTTCTGCTTTTAACCACTCCCTGGATGCATTTTTCCCTCC 1452
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTTGTTGCCCTGTGATTAGTTCTGCTTTTAACCACTCCCTGGATGCATTTTTCCCTCC 60

Qy       1453 TTGCATTCCCTCTTTTCCTGGAGTTCATACTAGAGAATCTGCACTATGTTTTTCCCTTT 1512
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGCATTCCCTCTTTTCCTGGAGTTCATACTAGAGAATCTGCACTATGTTTTTCCCTTT 120

Qy       1513 TTGTCTTGAGATGAAAGTTTTAAAATAATCCACCTCGTCATTCCACTCTCTGAACATC 1572
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTGTCTTGAGATGAAAGTTTTAAAATAATCCACCTCGTCATTCCACTCTCTGAACATC 180

Qy       1573 CCAAGCTGTATCCCTGGCCTCTTTTCCAGACTATGTTCTTTACTTGGGACCTAGAACT 1632
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAGCTGTATCCCTGGCCTCTTTTCCAGACTATGTTCTTTACTTGGGACCTAGAACT 240

RESULT 2
US-09-566-921-125
; Sequence 125, Application US/09566921
; Patent No. 6682888
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.
; APPLICANT: Tingley, Debora W.
```

APPLICANT: Edwards, Carla M.
TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE
FILE REFERENCE: PA-0024 US
CURRENT APPLICATION NUMBER: US/09/566,921
CURRENT FILING DATE: 2000-05-05
NUMBER OF SEQ ID NOS: 138
SOFTWARE: PERL Program
SEQ ID NO 125
LENGTH: 2494
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6682888 431588.3
NAME/KEY: unsure
LOCATION: 2018-2291, 2487
OTHER INFORMATION: a, t, c, g, or other
US-09-566-921-125

Query Match 7.9%; Score 196.8; DB 4; Length 2494;
Best Local Similarity 63.1%; Pred. No. 4.6e-38;
Matches 303; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

```
Qy   815 TTAATATCTTGCTTATGTCTACCTTACAGTTTAATCCCATGGAATCAAATTAAAT   874
             ||||||||||||||  |||      |||||||||  |||
Db   972 TTAAAAGGTTGTATGCCAGACAGATATCAGTTTAATTCCGTAAACCAATTACACCTGAG  1031

Qy   875 CATCATGACTACATTGATTCCCCATCGCTGAAGGACAGAATCATTGTGGCATTGTA    934
             |||  | |  |||       |||   |||||||||| |  |  |||||| | ||
Db  1032 CATTCTACTTTTATCACCTTCCATCTCTGAAGGACACAGAGATTCACTGTGTGGCTTATGTC 1091

Qy   935 TTTGATGCCAGCTCTATTCAATACTTCTCCCTCCAGATGATAGTAAAGATCAAAAGAATT   994
             || || | || ||||  || |  ||||   ||    | |  | ||| | ||||| ||||
Db  1092 TTAGACATCAACTCATTGACAATCTCTACTCTCTAAAATGTTGGCAAAAGTGAAGCAAGTT  1151

Qy   995 CGAAGGGAGTTGGTAAACGTGGTGTGGTACATGTGGCTTGCTCACTCATGTGGATAGC   1054
             |  | | | ||| |||||  |||| |  |||||||||  ||||||||  |||||||
Db  1152 CACAAAGAAGTATTAAACTGTGGTATAGCATATGTGGCCTTGCTTACTAAAGTGGATGAT   1211

Qy  1055 ATGGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAATGTGAGCCTGTGAGGTCC   1114
             |||||||||||||  ||| | |||  ||||| | ||||| | |||   | |||  |||
Db  1212 TGCAGTGAGGTTCTTCAAGACAACTTTAAACATGAGTAGATCTATGACTTCTCAAAGC   1271

Qy  1115 AAGCTAGAGAGGAAGTCCAAAGAAAAACTTGGATTGCTCTCTTCTGACATCTCGGTGGTTAGC  1174
             || |  |||| ||||||  | |  || ||||| || |||||||  |  ||| | ||||  ||
Db  1272 CGGGTCATGAAGTCCATAAAAATGCTAGGCATTCCTATTTCCAATATTTTGATGGTTGGA   1331

Qy  1175 AATTATTCCTCGAGTGGGGAGCTGACACCCTGTAAAGGATGTTCTCAAATTCTTTCTGCTCTG  1234
             |||||  || |  | ||| ||||| ||||| |||||  | || |||||||  ||| | |||||
Db  1332 AATTATGCTTCAGATTTGGAACTGGACCCCATGAAGGATATATTCTCATCCTTCATCCTCTGCACTG  1391

Qy  1235 AGACGAATGCTATGGGCTGCAGATGACTTCTTAGGAGGATTTGCCTTTTGAGCAAATAGGT   1294
             || | ||||||  ||||||| |||  ||| |||| |  ||||||| |  |||| || ||
Db  1392 AGGCAGATGCTGCCGGGCTGCGGATGCATGATTTTTTAGAGAAGATTTGCCTCTTCGAGGAAACTGGT  1451
```

RESULT 3
US-08-232-463-14, Application US/08232463
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 3.6%; Score 90.6; DB 1; Length 7218;
Best Local Similarity 14.4%; Pred. No. 2.8e-12;
Matches 81; Conservative 249; Mismatches 233; Indels 0; Gaps 0;

```
Qy  1374 GCTCTCTGTCTTTTGTGGGTTGTGCCGTGATTAGTTCTGCTTTTAACCCACTCCC   1433
             |:::: :::::::  :::  :: :: :::  :::: :::  ::::::::::::  |
Db  1064 GATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY   1123

Qy  1434 TGGATGCATTTTCCCTCCTTGCATTTCCCTCTTTTCCTGGAGTTCATACTAGAGAATCT   1493
             :  ::::::::::::::::::::::::  :: :  ::::::::: :: :::  ::::::
Db  1124 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY   1183

Qy  1494 GCACTATGTTTTTCCCTTTTGTCTTGAGATGAAAGTTTAAAATAATCCACCTCGTCA   1553
             : :::: ::::: :::::: :::: :::::::::: :: :::: ::::::::::::::
Db  1184 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY   1243

Qy  1554 TTTCCACTCTCTGAACATCCCAAGCTGTATCCCTGGCCTCTTTTCTCAGACTATGTTTCT   1613
             :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db  1244 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY   1303

Qy  1614 TTACTTGGGACCTAGACTGGATTGGCATTGCTCCTGATCAGATGAGACCTTTGA   1673
             ::::::::::::::: ::::::::::::::::::::::::::::::::: :: ::
Db  1304 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY   1363

Qy  1674 TTATTGCCCCTTCCTTAGGACCTTACACTCCTGTCTTCTTCTTGACTTGCCTTTTGTTT   1733
             :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db  1364 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY   1423

Qy  1734 CTTTCCTTCATCTTAGTCCCTCCTTCATCGCAGTAGCAGTTGCCAGGTAGAGGTATGTCC   1793
             :::::::::::::::::::::  || |||||||||||||| ||||||||||||||||||
Db  1424 YYYYYYYYYYGTACCAAATTCTTCTATCTTTAACTACTTGCATAGATAGGTAATTAC   1483

Qy  1794 TTTTATGTAATGCCACCGCATTAGTATTACATAAACTTTCTTTTAACAATCTGTGCAT   1853
             ||  |  |  || ||| || |   |   | ||| |||| |  |||| |  |  |||
Db  1484 AGTGATGCCTACATGCCGTGTTTTGAAACTGAAACTAGATTTGACAGAGGCGTCACGC   1543

Qy  1854 AGTACATGCTGCTCTGTTCCATTAGAGATTTGACAGAGGTTCAGTTAGTATACTCAA   1913
             || ||||| | | | | |||| || | | |||  |  |||  | ||| ||||| ||
Db  1544 TAGTCACAATCACCACTTTCATATTTAGAAATATAGATATGTAAAAATATAGTAGAATTTC   1603

Qy  1914 ATCTTATTTAGTGCTTGGGAAA   1936
```

```
Db    1604 ATTTTGTTTTTTCTATGCTATA 1626
              |||| |||| |  || |||

RESULT 4
US-09-385-982-122
; Sequence 122, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;   APPLICANT: ENDEGE, WILSON O., ET AL.
;   TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;   TITLE OF INVENTION: PRODUCTS: II
;   FILE REFERENCE: CCDNA-260XX
;   CURRENT APPLICATION NUMBER: US/09/385,982
;   CURRENT FILING DATE: 1999-08-30
;   EARLIER APPLICATION NUMBER: 09/328,111
;   EARLIER FILING DATE: 1999-06-08
;   EARLIER APPLICATION NUMBER: 60/117,393
;   EARLIER FILING DATE: 1999-01-27
;   EARLIER APPLICATION NUMBER: 60/098,639
;   EARLIER FILING DATE: 1998-08-31
;   NUMBER OF SEQ ID NOS: 544
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 122
;   LENGTH: 551
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(551)
;   OTHER INFORMATION: n = A,T,C or G
US-09-385-982-122

Query Match             2.2%;  Score 53.6;  DB 3;  Length 551;
Best Local Similarity  80.0%;  Pred. No. 0.0098;  Indels  4;  Gaps  3;
  Matches  96;  Conservative  0;  Mismatches  20;

Qy   845 TTAATTCCCATGGAATC-AATCAAATTAAATCATCATGACTACATTGATTCCCCATCGCT 903
             |||||||||||| | |||||||||||||||||||| ||||||||||||||||||
Db   427 TTAATTCCCATGGAATCAAATCAAATTAAATCATCATGACTACCTTGGTTCCCCATCGGT 486

Qy   904 GAA--GGACAGAATTCATTGTGTGGCATTTGTATTGAT-GCCAGCTCTATTCAATACTT 960
              ||  |||||||||||| |||||||||||||||||| ||||||||||||||||||||
Db   487 TGAAGGGACNCGNATTCATTGGGGNGGCATTGGATTTGATNNCNAGNTTTATTCAATACTT 546
```

RESULT 5
US-10-204-708-49/c
; Sequence 49, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;   TITLE OF INVENTION: by Assessing DNA Methylation
;   FILE REFERENCE: 5013.1012
;   CURRENT APPLICATION NUMBER: US/10/204,708
;   CURRENT FILING DATE: 2003-05-06
;   PRIOR APPLICATION NUMBER: PCT/EP01/03971
;   PRIOR FILING DATE: 2001-04-06
;   PRIOR APPLICATION NUMBER: DE 10019058.8
;   PRIOR FILING DATE: 2000-04-06
;   PRIOR APPLICATION NUMBER: DE 10019173.8
;   PRIOR FILING DATE: 2000-04-07
;   PRIOR APPLICATION NUMBER: DE 10032529.7
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: DE 10043826.1
;   PRIOR FILING DATE: 2000-09-01
;   NUMBER OF SEQ ID NOS: 98
;   SEQ ID NO 49
;   LENGTH: 6306
;   TYPE: DNA

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-49

```
Query Match             1.9%;  Score 48;  DB 4;  Length 6306;
Best Local Similarity  49.2%;  Pred. No. 0.05;
  Matches 126;  Conservative  0;  Mismatches 130;  Indels  0;  Gaps  0;

Qy     5 ATTCTTGTATTATAACCATAACAGTTCACTAATTAAATTAAATTAGGAATTGAATTGTT 64
           | |||||       |||          |||  ||||||||| ||||||||||||||||
Db   450 AATTTTAACTTTTAACTTTTAATTATTCCTAAATACCTTATTTTTATTACTTATTATAAAT 391

Qy    65 AAGTTAATTTGGTTTTATATTATTATGTTTAGCATTTATGTGGTTCAAAGATCAAATCTACA 124
           |||||||||   |||||| ||| |||||    |||||||||| ||||||||||||||||
Db   390 AATATCATTTTTTAAATAAAAATTTAACTAATTATTTTAATATAAAAATACCATTT 331

Qy   125 AAATAATGTATAGTCAAAGAATCTATCTTCCTTCCTCTGCCCCTCAAATAAATTCTCCC 184
           || |||| |||   ||  |||||||  ||||| ||| |  |||||||||||||||||||
Db   330 TAAAAATTTCTATCGATCAATTTTTTTAAAAATTTAAATAATAATTAAATAATTA 271

Qy   185 CTCTTCCCATTAGTAACCATATAAAATTATATTTACTTGCCTTTAAAATATGTAACA 244
           ||||| |||||||||||||||||| |||| ||||||||||||||||||||||||||||||
Db   270 TATTTCTTAAAAATTCTTACATATTTATTTTATTCTTCTTACCTTTTACCTAAAATAAAA 211

Qy   245 AAGTACATACATAAATTT 260
           ||| ||||||||||||
Db   210 CCTTCAATACCATTTT 195
```

RESULT 6
US-09-801-861-3
; Sequence 3, Application US/09801861
; Patent No. 6492154
; GENERAL INFORMATION:
;   APPLICANT: YAN, Chunhua et al.
;   TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;   TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;   TITLE OF INVENTION: THEREOF
;   FILE REFERENCE: CL001098
;   CURRENT APPLICATION NUMBER: US/09/801,861
;   CURRENT FILING DATE: 2001-03-09
;   NUMBER OF SEQ ID NOS: 10
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 53332
;   TYPE: DNA
;   ORGANISM: Human
US-09-801-861-3

```
Query Match             1.9%;  Score 47.8;  DB 4;  Length 53332;
Best Local Similarity  44.9%;  Pred. No. 0.12;
  Matches 181;  Conservative  0;  Mismatches 222;  Indels  0;  Gaps  0;

Qy     6 TCCTTGTATTATAACCATAACAGTTCACTAATTAAATTAAATTAGGAATTGAATTGTTA 65
           | |||||       |||          |||  ||||||||| ||||||||||||||||
Db 31258 TTGCTTTATCATCTCACATATATTTTTAATATTTATATATATTCTAAAAATGGCAATCACGTATATA 31317

Qy    66 AGTTAATTTGGTTTTATATTATTATGTTTAGCATTTATGTGGTTCAAAGATCAAATCTACAA 125
           ||||||||   |||||| ||| |||||    |||||||||| |||||||||||||||||
Db 31318 TTTTTATATATATTATATTTTATATATTTTATATAGTTATATATTTATAT 31377

Qy   126 AATAATGTATAGTCAAAGAATCTATCTTCCTTCCTCTGCCCCTCAAATAAATTCTCCCC 185
           || |||| |||   ||  |||||||  ||||| ||| |  |||||||||||||||||||
Db 31378 TTTATATATTATATTTTATAATATGTAACAA 31437

Qy   186 TCTTCCCATTAGTAACCATATAAAATTATATTTACTTGCCTTTAAAATATGTAACAA 245
           ||||| |||||||||||||||||| |||| ||||||||||||||||||||||||||||||
Db 31438 TATTTATTATATTTATATTTTATATTTTTATATATTTTTATATATATTTA 31497

Qy   246 AGTACATATATATATTGCTGCTACTCCCTTCTTAGAGAAGTGGTAGAAAACTATGTTATAT 305
           ||| ||||||||||||
Db 31498 TATATATATTTATATATATATTTTTATATATTTATATATTTATATATATTTATATAT 31557
```

Qy   306 TGACTTATCAGACATTGTTAACTGACATGGCATTTTCTGCTACAAATGTTCCAGCAGT 365
                ||||||| ||||| |||||
Db 31558 TTATATATTATATATATATTTATATATTTATATATATTTTATATATTTATATATAT 31617

Qy   366 TAATAAATCTTTGCATATATCATTTTGCATTTTTGTCAGTAGTAT 408
                ||||| |||| |||||
Db 31618 TTATATATTATATATATTTTTTTATATATATATATATATGTATTT 31660

RESULT 7
US-09-601-198-37
; Sequence 37, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
;   APPLICANT: Cassell, Gail H.
;   APPLICANT: Chen, Elson Y.
;   APPLICANT: Glass, Jennifer S.
;   APPLICANT: Glass, John I.
;   APPLICANT: Heiner, Cheryl R.
;   APPLICANT: Lefkowitz, Elliot
;   TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
;   TITLE OF INVENTION: UREALYTICUM
;   FILE REFERENCE: UAB-13452/22
;   CURRENT APPLICATION NUMBER: US/09/601,198
;   CURRENT FILING DATE: 2000-12-08
;   PRIOR APPLICATION NUMBER: 60/073,189
;   PRIOR FILING DATE: 1998-01-30
;   NUMBER OF SEQ ID NOS: 181
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
;   LENGTH: 1353
;   TYPE: DNA
;   ORGANISM: Ureaplasma urealyticum
US-09-601-198-37

Query Match   1.8%; Score 45.8; DB 4; Length 1353;
Best Local Similarity 48.6%; Pred. No. 0.1;
Matches 125; Conservative   0; Mismatches 132; Indels   0; Gaps   0;

Qy   7 TCTTGTATTATAACCATAACAGTTCACTAATTAAATTAGGAATTGAATTGTTAA 66
           ||||| ||||| ||||| ||||||| |||||||| |||||| ||
Db   964 TCGTTATTTCATCATTATAGAATTCCAGCTGAAAAAATATTTATAATATCAAAATAATA 1023

Qy   67 GTTAATTTGGTTTTATATTTTAGCATTTATGTGGTTCAAAGATCAAATCTACAAA 126
                ||||| |||| |||| ||||| ||||||| ||||||||||||
Db   1024 TGATATGAAATCATAAATAATTTTAAATTTATATTTTATTATTATTATTAACAAT 1083

Qy   127 ATAATGTATAGTCAAAGAATCTATCTTCCTTCCTTCGCCCTTCAAATAAATTTCTCCCCT 186
           ||||||| |||||||||||| |||||||||||||
Db   1084 ATCATCCTATTCAATGATCATCATAATTTTTTAAACTCTCTTTTAAATCTTTATCATC 1143

Qy   187 CTTCCCATTAGTAACCATATAAAAATTTATATTTTACTTGCCTTTTAAAATATGTAACAAA 246
           ||||| |||||||| ||| ||||||| |||||
Db   1144 GTTTATTTCTTTACCATTAAATAAAATTATAAAAACCAATAATTCATTTGGACT 1203

Qy   247 GTACATATAAATTTGCT 263
           |||| |||| || |||
Db   1204 TCATCCATAACTATGGT 1220

RESULT 8
US-07-991-867B-32/c
; Sequence 32, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
;   APPLICANT: Moyer, Richard W.
;   APPLICANT: Hall, Richard L.
;   APPLICANT: Gruidl, Michael E.
;   TITLE OF INVENTION: No. 547678lel Entomopoxvirus Expression System
;   NUMBER OF SEQUENCES: 66
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: David R. Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/07/991,867B
  FILING DATE: 12-DEC-1992
  CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: WO 92/14818
  FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/827,685
  FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/657,584
  FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
  NAME: Saliwanchik, David R.
  REGISTRATION NUMBER: 31,794
  REFERENCE/DOCKET NUMBER: UF114.C3
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 904-375-8100
  TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 32:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 660 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: unknown
  MOLECULE TYPE: DNA (genomic)
US-07-991-867B-32

Query Match   1.8%; Score 45.2; DB 1; Length 660;
Best Local Similarity 46.3%; Pred. No. 0.11;
Matches 189; Conservative   0; Mismatches 213; Indels   6; Gaps   1;

Qy   9 TTGTATTATACCATAACAGTTCACTAATTAAATTAGGAATTGAATTGTTAAGT 68
           ||||| |||| ||||||| ||||||| |||||| |||||
Db   436 TAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATTAATTTTTTTAAAT 377

Qy   69 TAATTTGGTTTTATATTTATGTTTAG------CATTTATGTGGTTCAAAGATCAAAATCTA 122
                ||| |||||| |||| |||| ||||||| |||| ||||||||
Db   376 TTACTAAATTATTAATAAAATATAGTCATTATAATATATGCATTATTATTACAATTAAAAATT 317

Qy   123 CAAAATAATGTATAGTCAAAGAATCTATCTTCCTTCCTTCGCCCCTTCAAATAAAATTTCTC 182
           ||||| |||| ||||| ||||||||||| ||||||||||
Db   316 CTATAGAATGTGGTAGTATATATTACTTATATATTGCTATTTTTGTATAGATATATCTA 257

Qy   183 CCCTCTTCCCATTAGTAACCATATAAAAATTTATATTTTACTTGCCTTTTAAAATATGTAA 242
           |||||| ||||||||| ||||| ||||||| ||||| ||||
Db   256 AATATGTATATTTTTTAATTTTGTTATAAAAATTAAATTAAAATTTAAATTTGAAA 197

Qy   243 CAAAGTACATACATATAAATTTGCTACTACCCCTTCTTAGGAATAATTATTTAAATATCATAATATATTA 302
           ||||| |||| |||| |||| |||| ||||||| |||||| |||| ||||
Db   196 TATATAAACTTTTAATATTTTCTGGAATATTTTCTGCTACAAATGTTCCAGC 137

Qy   303 TATTGACTTATCAGACATTGTTAACTGACATGGCATTTTCTGCTACAAATGTTCCAGC 362
           |||||||| |||| |||||| |||||| ||||||||| ||||||||||||||
Db   136 TATGCAATTCTCTAAATTAACTAATTTTTTTAATATATTAATATTAATAACATTTATCTC 77

Qy   363 AGTTAATAAATCTTGCATATATCATTTTGCATTTTGTCAGTATATCA 410
           ||| ||||||||| |||||| ||||||||| ||||||||
Db   76 TGTTTATTATTATTTTTTTTAAATTATAAATATTTTTAAAATATTTATTA 29

RESULT 9

Left column:

US-08-107-755A-32/c
; Sequence 32, Application US/08107755A
; Patent No. 5721352
; GENERAL INFORMATION:
;   APPLICANT: Moyer, Richard W.
;   APPLICANT: Hall, Richard L.
;   APPLICANT: Gruidl, Michael E.
;   TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
;   NUMBER OF SEQUENCES: 40
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: David R. Saliwanchik
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: Florida
;   COUNTRY: U.S.A.
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/107,755A
;   FILING DATE: 19-AUG-1993
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/827,658
;   FILING DATE: 30-JAN-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/657,584
;   FILING DATE: 19-FEB-1991
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Saliwanchik, David R.
;   REGISTRATION NUMBER: 31,794
;   REFERENCE/DOCKET NUMBER: UF114.C2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 32:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 660 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: unknown
;   MOLECULE TYPE: DNA (genomic)
US-08-107-755A-32

Query Match   1.8%; Score 45.2; DB 1; Length 660;
Best Local Similarity 46.3%; Pred. No. 0.11;
Matches 189; Conservative 0; Mismatches 213; Indels 6; Gaps 1;

```
Qy     9 TTGTATTATAACCATACAGTTCACTAATTAAATTAAATTTAGGAATTGAATTGTTAAGT    68
Db   436 TAGGAAAAACATTATTAAGTTACCAAATTTATTTTTAGATATTATTAATTTTTTTAAAT   377

Qy    69 TAATTTGGTTTTATATTTATGTTTAG------CATTTATGTGGTTCAAAGATCAAATCTA  122
Db   376 TTACTAAATTATTAAAATATAGTCATTTAATTAATTATTATTTTTT...             317

Qy   123 CAAAATAATGTATGTCAAAGAATCTATCTTCCTTCTCTGCCCCTTCAAATAAATTCTC    182
Db   316 CTATAGAATGTCGTAGTAGTAATATTATTACTTATTGCTATTTGTTATATAAGATATATCTA 257

Qy   183 CCCTCTTCCCATTAGTAACCATATAAAATTTATATTTTACTTGCCTTTAAAATATGTAA    242
Db   256 AATATGTTATATTTTTTAATTTTGTTATAAAAATTAAATTAATAATAATATTTAAATTTGAAA 197

Qy   243 CAAAGTACATATAAAATTGCTGCTACTCCCTTCTTAGAGAGTGGTAGAAAACTATGTTA    302
Db   196 TATATAAACTTTTAAATATTTTCTGGAATATTATTTAAAATATTATTATCATAATTATTA   137

Qy   303 TATTGACTTATCAGACATTGTTTAACTGACATGGCATTTTCTGCTACAAATGTTCCAGC    362
```

Right column:

```
Db   136 TATGCAATTCTTCAAATAACTAATTTTTTAATATATTAATTAATATTAATAACATTATCTC  77
Qy   363 AGTAATAATCTTGCATATACATTTGCATTTTGTCAGTATATCA                   410
Db    76 TGTTTATTATTATTTTTTTTAAATTATATATATTTAAAATATTTATTA               29
```

RESULT 10
US-08-544-332-32/c
; Sequence 32, Application US/0854332
; Patent No. 5935777
; GENERAL INFORMATION:
;   APPLICANT: Moyer, Richard W.
;   APPLICANT: Hall, Richard L.
;   APPLICANT: Gruidl, Michael E.
;   TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
;   NUMBER OF SEQUENCES: 77
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Gerard H. Bencen
;   STREET: 2421 N.W. 41st Street, Suite A-1
;   CITY: Gainesville
;   STATE: FL
;   COUNTRY: USA
;   ZIP: 32606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/544,332
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/991,867
;   FILING DATE: 07-DEC-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/107,755
;   FILING DATE: 19-AUG-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: WO 92/14818
;   FILING DATE: 12-FEB-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/827,685
;   FILING DATE: 30-JAN-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/657,584
;   FILING DATE: 19-FEB-1991
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Bencen, Gerard H.
;   REGISTRATION NUMBER: 35,746
;   REFERENCE/DOCKET NUMBER: UF114.C4
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 904-375-8100
;   TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 32:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 660 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: unknown
;   MOLECULE TYPE: DNA (genomic)
US-08-544-332-32

Query Match   1.8%; Score 45.2; DB 2; Length 660;
Best Local Similarity 46.3%; Pred. No. 0.11;
Matches 189; Conservative 0; Mismatches 213; Indels 6; Gaps 1;

```
Qy     9 TTGTATTATAACCATACAGTTCACTAATTAAATTAAATTTAGGAATTGAATTGTTAAGT    68
Db   436 TAGGAAAAACATTATTAAGTTACCAAATTTATTTTTAGATATTATTATTAATTTTTTTAAAT 377

Qy    69 TAATTTGGTTTTATATTTATGTTTAG------CATTTATGTGGTTCAAAGATCAAATCTA  122
```

```
Db  376  TTACTAAATTATTAATAAAAATTATAGTCATTATTAGTCATTATTACATGATTGATTCACCAATTTAAAAATT 317
              ||| ||||| |  ||   ||| ||||| |     |||  || ||  |||||    || ||
Qy  123  CAAAATAATGTATAGTCAAAGAATCATCTTCCTTCCTCTCTGCCCCTCAAATAAATTTCTC 182
              |||||||||||||||||||  ||||| |||  || |||
Db  316  CTATAGAATGTGGTAGTATATATTATTACTTATATTGCTATTTTGTATAGATATATCTA 257
              |||||||||||||  |||||||||||||||| ||  |||||   |||||||||
Qy  183  CCCTCTTCCCATTAGTAACCATATAAAATTTATATTTACTTGCCTTTTAAAATATGTAA 242
              ||||||||||||||| |||||
Db  256  AATATGTTATATTTTTAATTTTGTTATAAAATTTAAATTAAATTTAAATTTGAAA 197
              |||||||||||||||||  ||||| || |||||| ||||| ||||| ||||||
Qy  243  CAAAGTACATATAAAATTGCTGCTACTCCCTTCTTAGAGAAGTGGTAGAAAACTATGTTA 302
              |  |||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  196  TATATAAACTTTTATATTTTCTGGAATATTATTTAAAATATTATCATAATATATTA 137
              || || |||||||||||||||  ||||||
Qy  303  TATTGACTTATCAGACACATTGTTAACTGACAATGGCATTTTTCTGCTACAAATGTTCCAGC 362
              |||||||||||||||||||||  ||||||||||| |||||||||||||||||||
Db  136  TATGCAATTCTTCTAAATAACTAATTTTTTAATATATTAATTAATAACATTATCTC 77
              |||||||||||||||||||||||
Qy  363  AGTTAATAATCTTGCATATATCATTTGCATTTTGTCTCAGTATATCA 410
              ||| |||||||||||||||||||||||
Db  76  TGTTTATTATTATTTTTTTTAAATTATAATATTTAAAAATATTTATTA 29
```

```
Query Match         1.8%;  Score 45.2;  DB 4;  Length 660;
Best Local Similarity  46.3%;  Pred. No. 0.11;
Matches  189;  Conservative  0;  Mismatches  213;  Indels  6;  Gaps  1;

Qy    9  TTGTATTATAACCATAACAGTTCACTAATTAAATTAAATTTAGGATTGAATTGTTAAGT 68
            | ||||| ||||  |||| |||||||   ||| |||| ||| || |||||||
Db  436  TAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATTATTTTTTTAAAT 377
            || |||| |||||| || ||||||||
Qy   69  TAATTTGGTTTTATATATTTTATGTTTAG------CATTTATGTGGTTCAAAGATCAAAATCTA 122
            || ||| ||||| ||| ||| | |||       |||||||||| ||||||||||||
Db  376  TTACTAAATTATTAATAAAAATTATAGTCATTATTAGTCATTATTACATGATTGATTCACCAATTTAAAAATT 317
            |||||||||| ||| |||||  ||||| | | || |||||
Qy  123  CAAAATAATGTATAGTCAAAGAATCATCTTCCTTCCTCTCTGCCCCTCAAATAAATTTCTC 182
            |||||||||||||||||||  ||||| |||  ||
Db  316  CTATAGAATGTGGTAGTATATATTATTACTTATATTGCTATTTTGTATAGATATATCTA 257
            ||||||||||||||  ||||||||||
Qy  183  CCCTCTTCCCATTAGTAACCATATAAAATTTATATTTACTTGCCTTTTAAAATATGTAA 242
            |||||||||||||||  |||||
Db  256  AATATGTTATATTTTTAATTTTGTTATAAAATTTAAATTAAATTTAAATTTGAAA 197
```

```
;    LOCATION:  852..1511
US-07-991-867B-8

Query Match      1.8%;    Score 45.2; DB 1; Length 1511;
Best Local Similarity  46.3%;   Pred. No. 0.15;
Matches 189;  Conservative  0;  Mismatches 213;  Indels  6;  Gaps  1;

Qy     9 TTGTATTATAACCATAACAGTTCACTAATTAAAATTAAAATTAGGAATTGAATTGTTAAGT 68
Db  1287 TAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATTAATTTTTTTAAAT 1228

Qy    69 TAATTTGGTTTATATTATGTTTAG------CATTTATGTGGTTCAAAGATCAAATCTA 122
Db  1227 TTACTAAATTATTAATAAATTATGTCATTTATTAGTCATTTATTATATTACCAATTTAAAAATT 1168

Qy   123 CAAAATAATGATAGTCAAAGAATCTATCTTCCTTCCTGCCCCTTCAAATAAATTTCTC 182
Db  1167 CTATAGAATGTGGTAGTATAATTTATATTGTTATAATTGCTATTTTTGTTATAAGATATATCTA 1108

Qy   183 CCCTCTTCCCATTAGTAACCATATAAAATTTATATTTACTTGGCCTTTTAAAATATGTAA 242
Db  1107 AATATGTTATATTTTTCTGGAATATTATTTAAAATAATATTTAAATTAAATTGAA 1048

Qy   243 CAAGTACATAATAAAATTGCTGCTACTCCCCTTCCTTAGAGAAGTGGTAGAAAACTATGTTA 302
Db  1047 TATATAAACTTTTAATATTTTTCTGGAATATTATTTAATATATTCATAATAATATTA 988

Qy   303 TATTGACTTATCAGACATTGTTAACTGACATGGCATTTTCTGCTACAAATGTTCCAGC 362
Db   987 TATGCAATTCTTCTAAATTAACTAATTTTTTATATATTTTTAATATTAATAACATTATCTC 928

Qy   363 AGTTAATAATCTTTGCATATATCATTTTGTCAGTATATCA 410
Db   927 TGTTTATTATTATTTTTTTTAAATATAATATTTTAAAAATATTTATTA 880
```

RESULT 13
US-08-107-755A-8/c
```
; Sequence 8, Application US/08107755A
; Patent No. 5721352
; GENERAL INFORMATION:
;    APPLICANT: Moyer, Richard W.
;    APPLICANT: Hall, Richard L.
;    APPLICANT: Gruidl, Michael E.
;    TITLE OF INVENTION: No. 5721352e1 Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES: 40
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: David R. Saliwanchik
;       STREET: 2421 N.W. 41st Street, Suite A-1
;       CITY: Gainesville
;       STATE: Florida
;       COUNTRY: U.S.A.
;       ZIP: 32606
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/107,755A
;       FILING DATE: 19-AUG-1993
;       CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 07/827,658
;       FILING DATE: 30-JAN-1992
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 07/657,584
;       FILING DATE: 19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Saliwanchik, David R.
;       REGISTRATION NUMBER: 31,794
;       REFERENCE/DOCKET NUMBER: UF114.C2
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:   (904) 375-8100
;       TELEFAX:     (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 1511 base pairs
;    TYPE: nucleic acid
;    STRANDEDNESS: double
;    TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
;    ORGANISM: Amsacta moorei entemopoxvirus
; FEATURE:
;    NAME/KEY: CDS
;    LOCATION: complement (18..218)
; FEATURE:
;    NAME/KEY: CDS
;    LOCATION: complement (234..782)
; FEATURE:
;    NAME/KEY: CDS
;    LOCATION: 852..1511
US-08-107-755A-8

Query Match      1.8%;    Score 45.2; DB 1; Length 1511;
Best Local Similarity  46.3%;   Pred. No. 0.15;
Matches 189;  Conservative  0;  Mismatches 213;  Indels  6;  Gaps  1;

Qy     9 TTGTATTATAACCATAACAGTTCACTAATTAAAATTAAAATTAGGAATTGAATTGTTAAGT 68
Db  1287 TAGGAAAAACATTATTAAAGTTACCAAATTTATTTTTAGATATTATTAATTTTTTTAAAT 1228

Qy    69 TAATTTGGTTTATATTATGTTTAG------CATTTATGTGGTTCAAAGATCAAATCTA 122
Db  1227 TTACTAAATTATTAATAAATTATGTCATTTATTAGTCATTTATTATATTACCAATTTAAAAATT 1168

Qy   123 CAAAATAATGATAGTCAAAGAATCTATCTTCCTTCCTGCCCCTTCAAATAAATTTCTC 182
Db  1167 CTATAGAATGTGGTAGTATAATTTATATTGTTATAATTGCTATTTTTGTTATAAGATATATCTA 1108

Qy   183 CCCTCTTCCATTAGTAACCATATAAAATTTATATTTACTTGCCTTTTAAAATATGTAA 242
Db  1107 AATATGTTATATTTTTCTGGAATATTATTTAAAATAATATTTAAATTAAATTGAA 1048

Qy   243 CAAGTACACATAATAAAATTGCTGCTACTCCCCTTCCTTAGAGAAGTGGTAGAAAACTATGTTA 302
Db  1047 TATATAAACTTTTAATATTTTCTGGAATATTATTTAATAAATAATCATAATAATATTA 988

Qy   303 TATTGACTTATCAGACATTGTTAACTGACATGGCATTTTCTGCTACAAATGTTCCAGC 362
Db   987 TATGCAATTCTTCTAAATTAACAATTTTTTTATATATTTTTAATATTAATAACATTATCTC 928

Qy   363 AGTTAATAATCTTTGCATATATCATTTTGTCAGTATATCA 410
Db   927 TGTTTATTATTATTTTTTTTTAAATATAATATTTTTAAAATATTTATTA 880
```

RESULT 14
US-08-544-332-8/c
```
; Sequence 8, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
;    APPLICANT: Moyer, Richard W.
;    APPLICANT: Hall, Richard L.
;    APPLICANT: Gruidl, Michael E.
;    TITLE OF INVENTION: No. 5935777e1 Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES: 77
;    CORRESPONDENCE ADDRESS:
;       ADDRESSE: Gerard H. Bencen
;       STREET: 2421 N.W. 41st Street, Suite A-1
;       CITY: Gainesville
;       STATE: FL
;       COUNTRY: USA
;       ZIP: 32606
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/544,332
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/991,867
FILING DATE: 07-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/107,755
FILING DATE: 19-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bencen, Gerard H.
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET NUMBER: UF114.C4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entemopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852..1511

US-08-544-332-8

Query Match   1.8%;   Score 45.2;   DB 2;   Length 1511;
Best Local Similarity 46.3%; Pred. No. 0.15; Indels 6; Gaps   1;
Matches 189; Conservative 0; Mismatches 213;

Qy   9 TTGTATTATAACCATAACAGTTCACTAATTAAATTAAATTAGGAATTGAATTGTTAAGT 68
           ||||||| ||| ||| ||||| |||||| |||||||||| | |||||||||
Db 1287 TAGGAAAAACATTATTAAAGTTACCAAATTTATTTTAGATATTATTAATTTTTTTAAAT 1228

Qy   69 TAATTTGGTTTATATTATATGTTAG------CATTTATGTGGTTCAAAGATCAAAATCTA 122
           ||||| ||||| ||||||| ||| ||||||||| |||| || ||||| |||
Db 1227 TTACTAAATTATTAATAAAATATAGTCATTATTATTATAGTCATTTATTACATTTAAAAAATT 1168

Qy   123 CAAAATAATGTATAGTCAAAGAATCTATCTTCCTTCCTGCCCCTTCAAATAAATTTCTC 182
           |||||| |||| ||||||||||| | ||||||||||||
Db 1167 CTATAGAATGTGGTAGTATAATTACTTATATTGCTATTTTGTTATAAGATATATCTA 1108

Qy   183 CCCTCTTCCCATTAGTAACCATATAAAATTATATTTACTTGCCTTTAAAATATGTAA 242
           |||||||||||||||||||| ||||| |||||||| ||||| || |||||||||||
Db 1107 AATATGTTATATTTTTAATTTTGTTATAAAATTAATAAATTAATATTCATATATATTA 1048

Qy   243 CAAAGTACATATAAATTGCTGCTACTCCCCTTCCTAGAGAAGTGGTAGAAAACTATGTTA 302
           |||| || |||||| ||| ||||||||||| |||||||| |||||||||| ||||
Db 1047 TATATAAACTTTAATATTTGTTATAAAATTAAATAATAATAATATTA 988

Qy   303 TATTGACTTATCAGACACATTGTTAACTGACATGGCATTTTTCTGCTACAAATGTTCCAGC 362
           ||| | |||| ||||||||||||||| ||||||||||| ||||||| |||||||||||
Db   987 TATGCAATTCTTCTAAATTAACAATTTTTAATATTAATATTAATAACATTATCTC 928

Qy   363 AGTTAATAATCTTTGCATATATCATTTGCATTTTGTCAGTATATCA 410
           ||||||||||||||||||||||| ||||| ||||||||| ||||||
Db   927 TGTTTATTATTATTTTTTTTTTTAAATTATAATAATATTTTAAAAATATTTATTA 880

RESULT 15
US-09-370-861A-8/c
; Sequence 8, Application US/09370861A
; Patent No. 6410221
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 641022lel Entomopoxvirus Expression System
FILE REFERENCE: UF114.C4.D1
CURRENT APPLICATION NUMBER: US/09/370,861A
CURRENT FILING DATE: 1999-08-09
PRIOR APPLICATION NUMBER: US 07/991,867
PRIOR FILING DATE: 1992-12-07
PRIOR APPLICATION NUMBER: US 08/107,755
PRIOR FILING DATE: 1993-08-19
PRIOR APPLICATION NUMBER: WO 92/14818
PRIOR FILING DATE: 1992-02-12
PRIOR APPLICATION NUMBER: US 07/827,685
PRIOR FILING DATE: 1992-01-30
PRIOR APPLICATION NUMBER: US 07/657,584
PRIOR FILING DATE: 1991-02-19
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 1511
TYPE: DNA
ORGANISM: Amsacta moorei entomopoxvirus
US-09-370-861A-8

Query Match   1.8%;   Score 45.2;   DB 4;   Length 1511;
Best Local Similarity 46.3%; Pred. No. 0.15; Indels 6; Gaps   1;
Matches 189; Conservative 0; Mismatches 213;

Qy   9 TTGTATTATAACCATAACAGTTCACTAATTAAATTAAATTAGGAATTGAATTGTTAAGT 68
           ||||||| ||| ||| ||||| |||||| |||||||||| | |||||||||
Db 1287 TAGGAAAAACATTATTAAAGTTACCAAATTTATTTTAGATATTATTAATTTTTTTAAAT 1228

Qy   69 TAATTTGGTTTATATTATATGTTAG------CATTTATGTGGTTCAAAGATCAAAATCTA 122
           ||||| ||||| ||||||| ||| ||||||||| |||| || ||||| |||
Db 1227 TTACTAAATTATTAATAAAATATAGTCATTATCATTTATTACATTTAAAAAATT 1168

Qy   123 CAAAATAATGTATAGTCAAAGAATCTATCTTCCTTCCTGCCCCTTCAAATAAATTTCTC 182
           |||||| |||| ||||||||||| | ||||||||||||
Db 1167 CTATAGAATGTGGTAGTATAATTACTTATATTGCTATTTTGTTATAAGATATATCTA 1108

Qy   183 CCCTCTTCCCATTAGTAACCATATAAAATTATATTTACTTGCCTTTAAAATATGTAA 242
           |||||||||||||||||||| ||||| |||||||| ||||| || |||||||||||
Db 1107 AATATGTTATATTTTTAATTTTGTTATAAAATTAATAAATTAATATTCATATATATTA 1048

Qy   243 CAAAGTACATATAAATTGCTGCTACTCCCCTTCCTAGAGAAGTGGTAGAAAACTATGTTA 302
           |||| || |||||| ||| ||||||||||| |||||||| |||||||||| ||||
Db 1047 TATATAAACTTTAATATTTGTTATAAAATTAAATAATAATAATATTA 988

Qy   303 TATTGACTTATCAGACACATTGTTAACTGACATGGCATTTTTCTGCTACAAATGTTCCAGC 362
           ||| | |||| ||||||||||||||| ||||||||||| ||||||| |||||||||||
Db   987 TATGCAATTCTTCTAAATTAACAATTTTTAATATTAATATTAATAACATTATCTC 928

Qy   363 AGTTAATAATCTTTGCATATATCATTTGCATTTTGTCAGTATATCA 410
           ||||||||||||||||||||||| ||||| ||||||||| ||||||
Db   927 TGTTTATTATTATTTTTTTTTTTAAATTATAATAATATTTTAAAAATATTTATTA 880

us-10-051-835-18.rni

```
                 GenCore version 5.1.6
         Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    April 24, 2004, 23:47:42 ; Search time 660.01 Seconds
                                      (without alignments)
                                      15988.422 Million cell updates/sec

Title:           US-10-051-835-18
Perfect score:   2484
Sequence:        1 gttattcctgattataac.........ataaattcttataaactc 2484

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:    6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_29Jan04:*
         1:  geneseqm1980s:*
         2:  geneseqm1990s:*
         3:  geneseqm2000s:*
         4:  geneseqm2001as:*
         5:  geneseqm2001bs:*
         6:  geneseqm2002s:*
         7:  geneseqm2003as:*
         8:  geneseqm2003bs:*
         9:  geneseqm2003cs:*
        10:  geneseqm2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2484	100.0	2484	7	ABX63692	Abx63692 Human cDN
2	1358	54.7	2741	8	ACD42248	Acd42248 Human hep
3	1358	54.7	2741	9	ADE25678	Ade25678 Human cDN
4	1235.2	49.7	2527	4	AAS60823	Aas60823 Human can
5	456.6	18.4	2068	2	ABX63519	Abx63519 Human cDN
6	455	18.3	1335	2	AAX33941	Aax33941 Human HCM
7	455	18.3	1679	6	ABN59995	Abn59995 Novel hum
8	448.6	18.1	1332	1	AAN81259	Aan81259 Sequence
9	415.4	16.7	567	4	AAS60407	Aas60407 Human can
10	368	14.8	482	6	ABL80943	Abl80943 Human cDN
11	354	14.3	2515	7	ABX63520	Abx63520 Human ova
12	331.2	13.3	580	4	AAS60100	Aas60100 Human can
13	315.4	12.7	977	8	AAH34468	Aah34468 Human col
14	311	12.5	3276	4	AAS27231	Aas27231 cDNA enco
15	311	12.5	3276	9	ADB93409	Adb93409 Human cDN
16	281	11.0	600	7	ABX63603	Abx63603 Human ade
c 17	272	11.0	272	6	ABK29513	Abk29513 Colon ade
18	240	9.7	240	6	ABT06900	Abt06900 Human ova
19	240	9.7	240	7	ABX72778	Abx72778 Human ova
20	196.8	7.9	2058	6	ABS76439	Abs76439 cDNA enco
21	196.8	7.9	2058	9	ADD18927	Add18927 Human dis
22	196.8	7.9	2062	3	AAA08592	Aaa08592 Human cyt
23	196.8	7.9	2062	7	ABX63510	Abx63510 Human cDN

	Score	%	Length	DB	ID	Description
c 24	163.4	6.6	652	6	ABQ57002	Human col
25	101.6	4.1	14160	4	AAK65583	Human imm
26	101.2	4.0	489	4	ABK44482	cDNA enco
c 27	100.6	4.0	499	4	AAL23095	Human bre
28	99.4	4.0	227	7	ABT14722	Human Bcl
29	98.4	3.8	135	7	ABT14802	Human Bcl
30	95.4	2.4	410	2	AAT22120	Human gen
31	60	2.4	60	6	ABN46205	Human spl
c 32	59.4	2.4	7490	6	ABL32283	Human imm
33	58.8	2.4	65359	6	AAD46665	Human tra
c 34	57.8	2.3	6075	6	ABL70596	Chemical1
c 35	57.8	2.3	6075	6	AAS61317	Human gen
36	54.6	2.2	15387	6	ABL32184	Human imm
37	54.4	2.2	337	3	AAC05156	Human sec
c 38	54.2	2.2	8056	7	ABZ10246	Haematopo
c 39	54.2	2.2	8056	7	ABZ10100	Haematopo
40	53.6	2.2	551	3	AAA16117	Human col
c 41	53	2.1	2000	7	ADA71938	Rice gene
42	53	2.1	159400	6	ABQ88126	Human ost
c 43	52	2.1	6071	6	ABL92214	Chemical1
c 44	52	2.1	6071	6	AAD22315	Chemical1
45	51.6	2.1	6106	4	AAS46429	Tumour su

ALIGNMENTS

```
RESULT 1
ABX63692
ID   ABX63692 standard; cDNA; 2484 BP.
XX
AC   ABX63692;
XX
DT   26-FEB-2003  (first entry)
XX
DE   Human cDNA #692 differentially expressed in activated vascular tissue.
XX
KW   Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;
KW   hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;
KW   gene therapy; vascular disease; cancer; coronary; artery disease;
KW   hypertension; diabetes; pre-eclampsia; restenosis;
KW   ischaemia-reperfusion injury; stroke.
XX
OS   Homo sapiens.
XX
PN   US2002137081-A1.
XX
PD   26-SEP-2002.
XX
PF   08-JAN-2002; 2002US-00044090.
XX
PR   28-JUL-2000; 2000US-0222469P.
PR   08-JAN-2001; 2001US-0260483P.
XX
PA   (BAND/) BANDMAN O.
XX
PI   Bandman O;
XX
DR   WPI; 2003-110597/10.
XX
PT   Combination for diagnosing, staging, treating, or monitoring the
PT   progression of treatment of a vascular disease, e.g. atherosclerosis,
PT   comprises several cDNAs that are differentially expressed in activated
PT   vascular tissue.
XX
PS   Claim 1; Page; 18pp; English.
XX
CC   This invention relates to a combination comprising several cDNAs that are
CC   differentially expressed in activated vascular tissue. The invention also
CC   discloses a high throughput method for detecting differentially expressed
CC   cDNAs in a sample. The cDNAs of the invention may have
CC   antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC   gynaecological; vasotropic and cerebroprotective activities and may be
```

```
CC  used in gene therapy. The cDNAs of the invention may be used in a high-
CC  throughput methods for detecting differential expression of one or more
CC  cDNAs in a sample. or screening several molecules or compounds to
CC  identify a molecule or compound that specifically binds a cDNA of the
CC  invention. A protein encoded by the cDNA may be used to screen several
CC  molecules or compounds to identify a ligand that specifically binds to
CC  the protein, or to produce or purify an antibody to the protein that can
CC  be used to detect a protein in a sample or purify a natural or
CC  recombinant protein from sample. The nucleotides may be useful for
CC  diagnosing, staging, treating, or monitoring the progression of treatment
CC  of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
CC  disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion
CC  injury, restenosis, or stroke. The cDNAs can also be used for large-scale
CC  genetic or gene expression analysis of several new nucleic acid
CC  molecules. Antibodies to the proteins encoded by the cDNAs are useful for
CC  diagnosing pre-pathologic disorders, and chronic or acute diseases
CC  associated with abnormalities in the expression, amount or distribution
CC  of the protein. The present sequence represents a cDNA of the invention
CC  that is differentially expressed in activated vascular tissue. Note: The
CC  sequence data for this patent did not form part of the specification, but
CC  was obtained in electronic format directly from USPTO at
CC  http.seqdata.uspto.gov/sequence.html?DocID=2002137081

XX
SQ  Sequence 2484 BP; 657 A; 488 C; 436 G; 903 T; 0 U; 0 Other;

Query Match       100.0%;   Score 2484;   DB 7;   Length 2484;
Best Local Similarity  100.0%;   Pred. No. 0;
Matches 2484;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy    1  GTTTATTCTTGTATTATAACCATAACAGTTCACTAATTAAATTAAATTAGGAATTGAAT  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  GTTTATTCTTGTATTATAACCATAACAGTTCACTAATTAAATTAAATTAGGAATTGAAT  60

Qy   61  TGTTAAGTTAATTTGGTTTATATTTATATATTGTTAGCATTTATGTTGGTTCAAAGATCAAATC  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  TGTTAAGTTAATTTGGTTTATATTTATATATTGTTAGCATTTATGTTGGTTCAAAGATCAAATC  120

Qy  121  TACAAAAATAATGTATAGTCAAAGAATCATCTTCCTTCTTGTCTCTGCCCCTTCAAATAAATTC  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  TACAAAAATAATGTATAGTCAAAGAATCATCTTCCTTCTTGTCTCTGCCCCTTCAAATAAATTC  180

Qy  181  TCCCCTCTTCCATTAGTAACCATATAAAATTATATATTTACTTGCCTTTAAAATATGT  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TCCCCTCTTCCATTAGTAACCATATAAAATTATATATTTACTTGCCTTTAAAATATGT  240

Qy  241  AACAAAGTACATATAAAATTGCTGTACTCCCTTCTTAGAGAAGTGGTAGAAAACTATGT  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  AACAAAGTACATATAAAATTGCTGTACTCCCTTCTTAGAGAAGTGGTAGAAAACTATGT  300

Qy  301  TATATTGACTTATCAGACATTGTCTGCTACTGACATGGCATTTTGTCTACAAATGTTCCA  360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  TATATTGACTTATCAGACATTGTCTGCTACTGACATGGCATTTTGTCTACAAATGTTCCA  360

Qy  361  GCAGTTAATAATCTTTGCATAATACATTTTGCATTTTGTCAGTATATCAGTGGGACAGA  420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  GCAGTTAATAATCTTTGCATAATACATTTTGCATTTTGTCAGTATATCAGTGGGACAGA  420

Qy  421  TTCCCAACAGAGAAGTGCTAGATGAAGAGTAAAGTCATCTCACCTGGACCCCTTCCCTC  480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  TTCCCAACAGAGAAGTGCTAGATGAAGAGTAAAGTCATCTCACCTGGACCCCTTCCCTC  480

Qy  481  CTTGCGCTCTGCTTGATGGTACCAGCTTTTTTTCTGAATGCTCTTATTCTCTTAGCAGACT  540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  CTTGCGCTCTGCTTGATGGTACCAGCTTTTTTTCTGAATGCTCTTATTCTCTTAGCAGACT  540

Qy  541  GCAGATTGCCTCTCTATCTAGAACACTTTCAGCCATATCCTTATTCTCTTAGCAGACT  600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  GCAGATTGCCTCTCTATCTAGAACACTTTCAGCCATATCCTTATTCTCTTAGCAGACT  600

Qy  601  CACTCCTGCTATTTAGCTCCAGATTACATGTCCTTCCTCCAGAGGCCTTACTTGAC  660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  CACTCCTGCTATTTAGCTCCAGATTACATGTCCTTCCTCCAGAGGCCTTACTTGAC  660

Qy  661  TAGGTGAAAGGACTGCTGGCTACTTGCATAATGCTGTCCTCTCCACCAAGATTGTTAG  720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661  TAGGTGAAAGGACTGCTGGCTACTTGCATAATGCTGTCCTCTCCACCAAGATTGTTAG  720

Qy  721  CCTCATACTATTAGAGACTTTCGTTTGTTTGTTGTTTACCATTTGATTCACTTGCACAGTGCCT  780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721  CCTCATACTATTAGAGACTTTCGTTTGTTTGTTGTTTACCATTTGATTCACTTGCACAGTGCCT  780

Qy  781  GGTACATAGTTTGTGTCATAAAATATTTTTGAATAATAATATCTTGCTTTATGTCTACCTT  840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781  GGTACATAGTTTGTGTCATAAAATATTTTTGAATAATAATATCTTGCTTTATGTCTACCTT  840

Qy  841  ACAGTTTAATCCCATGAAATCAATCAAATTAAATCATCATGACTACATTGATTCCCCATC  900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841  ACAGTTTAATCCCATGAAATCAATCAAATTAAATCATCATGACTACATTGATTCCCCATC  900

Qy  901  GCTGAAGGACAGAATTCATTGTGTGCATTTGTATTTGATGCCAGTCTATTCAATACTT  960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901  GCTGAAGGACAGAATTCATTGTGTGCATTTGTATTTGATGCCAGTCTATTCAATACTT  960

Qy  961  CTCCTCCTCAGATGATAGTAAAGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGTGT  1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961  CTCCTCCTCAGATGATAGTAAAGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGTGT  1020

Qy 1021  GGTACATGTGGCTTGCTCACTCATGTGGATAGCATGGATTGATTACAAAAGGTGACCT  1080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021  GGTACATGTGGCTTGCTCACTCATGTGGATAGCATGGATTGATTACAAAAGGTGACCT  1080

Qy 1081  TATAGAAATAGAGATGTGAGCCTGTGAGGTCCAAGCTAGAGGAGTCCAAAGAAAAACT  1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081  TATAGAAATAGAGATGTGAGCCTGTGAGGTCCAAGCTAGAGGAGTCCAAAGAAAAACT  1140

Qy 1141  TGGATTTGCTCTTCTGACATCTCGGTGGTTAGCAATTATCCTCTGAGTGGGAGCTGGA  1200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141  TGGATTTGCTCTTCTGACATCTCGGTGGTTAGCAATTATCCTCTGAGTGGGAGCTGGA  1200

Qy 1201  CCCTGTAAAGATGTTCTAATTCTTCTGCTCTGAGACGAATGCTATGGCTGCAGATGA  1260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201  CCCTGTAAAGATGTTCTAATTCTTCTGCTCTGAGACGAATGCTATGGCTGCAGATGA  1260

Qy 1261  CTTCTTAGAGGATTGCCTTTGAGCAAATAGGTAGTGATGGTTGGTGTGGAAGCTTG  1320
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261  CTTCTTAGAGGATTGCCTTTGAGCAAATAGGTAGTGATGGTTGGTGTGGAAGCTTG  1320

Qy 1321  GAAGCGGTCAGGTAGTTGGCTACTTTCTCGGATCTATTAAATACCTGGCAGCTCTCT  1380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321  GAAGCGGTCAGGTAGTTGGCTACTTTCTCGGATCTATTAAATACCTGGCAGCTCTCT  1380

Qy 1381  GTCTTTTTGTGGGTGTTGCCCTCTCTTTTCCTGGAGTTCATACTAGAGAATCTGCACTAT  1440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381  GTCTTTTTGTGGGTGTTGCCCTCTCTTTTCCTGGAGTTCATACTAGAGAATCTGCACTAT  1440

Qy 1441  ATTTTCCCTCCTTGCATTCCCTCTTTTCCTGGAGTTCATACTAGAGAATCTGCACTAT  1500
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441  ATTTTCCCTCCTTGCATTCCCTCTTTTCCTGGAGTTCATACTAGAGAATCTGCACTAT  1500

Qy 1501  GTTTTTCCCTTTTGTCTTGGATGAAAGTTTAAAATAATCCACTCGTCATTCCAC  1560
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501  GTTTTTCCCTTTTGTCTTGGATGAAAGTTTAAAATAATCCACTCGTCATTCCAC  1560

Qy 1561  TCTCTGAACATCCAAGCTGTATCCCTGGCGCCTCTTTCTCGACTATGTTTCTTTACTTG  1620
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561  TCTCTGAACATCCAAGCTGTATCCCTGGCGCCTCTTTCTCGACTATGTTTCTTTACTTG  1620

Qy 1621  GGACCTAGAACTGGATTGGAGCATTGGCATTGCTCCTGATCAGATGAGACCTTTGATTATTTG  1680
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621  GGACCTAGAACTGGATTGGAGCATTGGCATTGCTCCTGATCAGATGAGACCTTTGATTATTTG  1680

Qy 1681  CCCCTTCCTTAGGACCTTACACTCCTGTCTTTGACTTGCCTTTTGTTTCTTTCCT  1740
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681  CCCCTTCCTTAGGACCTTACACTCCTGTCTTTGACTTGCCTTTTGTTTCTTTCCT  1740

Qy 1741  TCATCTTAGTCCCTCTCATGCAGTATGGTCATTGCTAGGTAGAGGTATGTCCTTTATG  1800
```

```
XX
PA   (LASE/)  LASEK A K W.
PA   (JONE/)  JONES D A.
PA   (KARP/)  KARPF A R.
XX
PI   Lasek AKW, Jones DA, Karpf AR;
XX
DR   WPI; 2003-503249/47.
XX
PT   New combination comprising cDNAs that are expressed in a disorder or
PT   process associated with DNA methylation, useful for diagnosing, staging,
PT   treating or monitoring treatment of cancer, e.g. colon cancer.
XX
PS   Claim 2; Page 58-59; 66pp; English.
XX
CC   The invention relates to a combination comprising cDNAs which are
CC   expressed in a disorder or process associated with DNA methylation. The
CC   combination and cDNAs are useful for diagnosing, staging, treating or
CC   monitoring treatment of cancer, e.g. colon cancer and for detecting
CC   changes in expression of genes encoding proteins that are associated with
CC   DNA methylation. The protein is useful for screening molecules or
CC   compounds to identify at least one ligand that binds to the protein and
CC   for producing an antibody. The present sequence represents a cDNA
CC   expressed in a disorder or process associated with DNA methylation
XX
SQ   Sequence 2741 BP; 841 A; 452 C; 529 G; 919 T; 0 U; 0 Other;

Query Match             54.7%;  Score 1358;  DB 8;  Length 2741;
Best Local Similarity   99.9%;  Pred. No. 5.6e-296;
Matches 1369;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  1;

Qy   1116 AGCTAGGAGGAGTCCAAAGAAAACTTGGATTTGCTCTTCTGACAATCTCGGTGGGTTAGCA 1175
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1364 AGCTAGGAGGAGTCCAAAGAAAACTTGGATTTGCTCTTCTGACAATCTCGGTGGGTTAGCA 1423

Qy   1176 ATTATTCCTCGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTCTGCTCTGA 1235
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1424 ATTATTCCTCGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTCTGCTCTGA 1483

Qy   1236 GACGAATGCTATGGGCTGCAGATGACTTCTAGAGGATTTGCCTTTTGGAGCAAATAGGTA 1295
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1484 GACGAATGCTATGGGCTGCAGATGACTTCTAGAGGATTTGCCTTTTGGAGCAAATAGGTA 1543

Qy   1296 GATG-GTTTGGTGGTGTGGAAGCTTGGAAGCCGTCAGGTAGTTGGCTACTTCTGCTTGG 1354
             |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1544 GATGTGTTTGGTGGTGTGGAAGCTTGGAAGCCGTCAGGTAGTTGGCTACTTCTGCTTGG 1603

Qy   1355 ATCTATTAAATACCTGGCAGCTCTCGTGCTCTTTTGTGGGTTGTTGCCCTGTGATTAGTTC 1414
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1604 ATCTATTAAATACCTGGCAGCTCTCGTGCTCTTTTGTGGGTTGTTGCCCTGTGATTAGTTC 1663

Qy   1415 TGCTTTTAACCCACTCCCTGGATGCATTTTCCCTCCTTTGTCTTGAGATGAAAGTTTTA 1474
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1664 TGCTTTTAACCCACTCCCTGGATGCATTTTCCCTCCTTTGTCTTGAGATGAAAGTTTTA 1723

Qy   1475 AGTTCATACTAGAGAATCTGCACTATGTTTCCTTTTGCCCTTTGCCCTTTGTCTTGGGCCTCT 1534
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1724 AGTTCATACTAGAGAATCTGCACTATGTTTCCTTTTGCCCTTTGCCCTTTGTCTTGGGCCTCT 1783

Qy   1535 AAATAATCCACCTCTGTCATTTCCACTCTCTGAACATCCAAGCTGTATCCCTGGCCCT 1594
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1784 AAATAATCCACCTCTGTCATTTCCACTCTCTGAACATCCAAGCTGTATCCCTGGCCCT 1843

Qy   1595 TTTTCGATGATGTTCTTTACTTGGGACCTAGAACTGGATTGGCATTGCTCCT 1654
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1844 TTTTCGATGATGTTCTTTACTTGGGACCTAGAACTGGATTGGCATTGCTCCT 1903

Qy   1655 GATCAGATGAGACCTTGATTATTTGCCCTTTCGTCTTTCT 1714
             ||||||||||||||||||||||||||||||||||||||||||
Db   1904 GATCAGATGAGACCTTGATTATTTGCCCTTTCGTCTTTCT 1963

Qy   1715 TTGACTTGCCTTTTGTTTCTTTCCTTCATCTTAGTCCCTTCTTCATGCAGTATGGTCATT 1774
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1741 TCATCTTAGTCCCTTCTTCATGCAGTATGTTCATTGCATTTGCTTATG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1801 TAATGGGCCACCGCATTTAGTATTACATAAAACTTCTTTTAACAATCTGTGCATAGTACAT 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TAATGGGCCACCGCATTTAGTATTACATAAAACTTCTTTTAACAATCTGTGCATAGTACAT 1860

Qy   1861 GCTGCTCTGTCATTAGAGATTGACAGAGGTTTCAGTTAGTATACTCAAATCTTAT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GCTGCTCTGTCATTAGAGATTGACAGAGGTTTCAGTTAGTATACTCAAATCTTAT 1920

Qy   1921 TTTAGTGCTTGGGAAATCAATTCAGAATATCACATCCTCTCAATTCTCTCTTACTCAAA 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TTTAGTGCTTGGGAAATCAATTCAGAATATCACATCCTCTCCAATTCTCTCTTACTCAAA 1980

Qy   1981 TTGCTGGGAAACTCATCATGTTACTAACTTGTTGTGCTCTAAACTCTGCCATCTGGTTTCCC 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 TTGCTGGGAAACTCATCATGTTACTAACTTGTTGTGCTCTAAACTCTGCCATCTGGTTTCCC 2040

Qy   2041 CATCCCTTCTTCTTCCTCATGGTACGTGTGCTCCTAATATTAGCGTTGGTTGAGATTTTCA 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 CATCCCTTCTTCTTCCTCATGGTACGTGTGCTCCTAATATTAGCGTTGGTTGAGATTTTCA 2100

Qy   2101 GTGGTCCAATATTCCTCTTCCCTCGGTTGCCTTTCCTGAGATAATCCACTAAGAATATT 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GTGGTCCAATATTCCTCTTCCCTCGGTTGCCTTTCCTGAGATAATCCACTAAGAATATT 2160

Qy   2161 TTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGAAATTATCAACTGTGCACAAGAGAAAA 2220
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGAAATTATCAACTGTGCACAAGAGAAAA 2220

Qy   2221 AATAGATATGTGAAAGGTTCACGTAAATTCCTCACATCACAGAAGATTAAAATTCAGAA 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 AATAGATATGTGAAAGGTTCACGTAAATTCCTCACATCACAGAAGATTAAAATTCAGAA 2280

Qy   2281 AGGAGAAAACACAGACCAAGAGAAGTATCAAGGACCAAAGAGTATCAAGGATGTGTTTATTAATGT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 AGGAGAAAACACAGACCAAGAGAAGTATCAAGGACCAAAGAGTATCAAGGATGTGTTTATTAATGT 2340

Qy   2341 CTAGGATGAAGAAATGCATAGAACATTGTAGTACTTGTAAATAACTAGAAATAACATGAT 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CTAGGATGAAGAAATGCATAGAACATTGTAGTACTTGTAAATAACTAGAAATAACATGAT 2400

Qy   2401 TTAGTCATAATTGTGAAAAATAATAATTTTCTTGGATTTATGTTCTGTATCTGTGA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TTAGTCATAATTGTGAAAAATAATAATTTTCTTGGATTTATGTTCTGTATCTGTGA 2460

Qy   2461 AAAAATAAATTTCTTATAAAACTC 2484
             ||||||||||||||||||||||||
Db   2461 AAAAATAAATTTCTTATAAAACTC 2484

RESULT 2
ACD42248
ID   ACD42248 standard; cDNA; 2741 BP.
XX
AC   ACD42248;
XX
DT   05-SEP-2003  (first entry)
XX
DE   Human hepatitis C microtubular aggregate cDNA Incyte 231486.26.
XX
KW   Human; ss; DNA methylation; cancer; colon cancer.
OS   Homo sapiens.
XX
PN   US2003013099-A1.
XX
PD   16-JAN-2003.
XX
PF   07-MAR-2002; 2002US-00093766.
XX
PR   19-MAR-2001; 2001US-0277380P.
```

XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Mikita T, Shiffman D, Porter JG, Kaser MR;
XX
DR   WPI; 2003-875398/81.
XX
PT   Combination containing several polynucleotide that are differentially
PT   expressed in foam cells and complements of the polynucleotides, useful
PT   for diagnosing cardiovascular disease or atherosclerosis.
XX
PS   Claim 1; SEQ ID NO 82; 37pp; English.
XX
CC   The invention relates to a combination comprising several polynucleotides
CC   having any one of 127 sequences (S1) such as the sequence of human
CC   calmodulin gene, human mRNA for KIAA0930 protein, leukotriene A4
CC   hydrolase, human CGI-142 protein mRNA, human K+ channel beat 2 subunit
CC   mRNA, etc., and their complements. The cDNAs are differentially expressed
CC   in LPS (lipopolysaccharide)-treated foam cells. Also included are
CC   obtaining an extended or full length gene from a library of nucleic acid
CC   sequences, an expression vector containing the nucleic acids, a host cell
CC   containing the vector, a purified polypeptide appearing as ADE25750 and
CC   ADE25751, producing a protein by culturing the host cell, and a
CC   composition comprising a purified antibody that specifically binds to the
CC   proteins. The foam cell-expressed nucleic acids are useful for a high
CC   throughput detection of differential expression of one or more
CC   polynucleotides in a sample. The sample is from a subject with
CC   atherosclerosis and comparison with a standard defines early, mid or late
CC   stages of the disorder. The foam cell-expressed nucleic acids are useful
CC   for high throughput screening of a library of molecules or compounds to
CC   identify a ligand which binds a polynucleotide. The library is chosen
CC   from DNA molecules, peptides, proteins and RNA molecules. The protein is
CC   useful for a high throughput screening of library of molecules or
CC   compounds to identify at least one ligand which specifically binds a
CC   protein, for purifying a ligand from a sample for making a antibody. The
CC   foam cell-expressed nucleic acids are useful for diagnosing
CC   cardiovascular disorder. The foam cell-expressed nucleic acids are useful
CC   as elements on a microarray which can be used for detecting related
CC   polynucleotide in a sample, diagnosing cardiovascular disease,
CC   atherosclerosis. The present sequence represents a cDNA whose expression
CC   is upregulated in LPS treated foam cells.
XX
SQ   Sequence 2741 BP; 841 A; 452 C; 530 G; 918 T; 0 U; 0 Other;

Query Match   54.7%; Score 1358; DB 9; Length 2741;
Best Local Similarity   99.9%; Pred. No. 5.6e-296;
Matches 1369; Conservative   0; Mismatches   0; Indels   1; Gaps   1;

```
Qy   1116 AGCTAGAGGAAGTCCAAAGAAAACTTGGATTGCTCTTTCTGACATCTCGGTGGTTAGCA 1175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1364 AGCTAGAGGAAGTCCAAAGAAAACTTGGATTGCTCTTTCTGACATCTCGGTGGTTAGCA 1423

Qy   1176 ATTATTCCTCTGAGTGGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 1235
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1424 ATTATTCCTCTGAGTGGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 1483

Qy   1236 GACGAATGCTATGGGCTGCAGATGACTTCCTTAGAGAGATTGCCTTTTGAGCAAATAGGTA 1295
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1484 GACGAATGCTATGGGCTGCAGATGACTTCCTTAGAGAGATTGCCTTTTGAGCAAATAGGTA 1543

Qy   1296 GATG-GTTTGGTGGTGTGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGG 1354
          ||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1544 GATGTGTTTGGTGGTGTGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGG 1603

Qy   1355 ATCTATTAAATAACCTGGCCAGCTCCTGTGCTTTTGTGGGTGTTGCCCGTGGTTAGTTC 1414
          |||||||||||||||||||||||||||||  ||||| ||||||||||||||| |||||
Db   1604 ATCTATTAAATAACCTGGCCAGCTCCTGTCTTTTTTGTGGGTGTTGCCCGTGATTAGTTC 1663

Qy   1415 TGCTTTTTAACCCACTCCCTGGATGCATTTTCCCTCCTTGCATTTCCCTCTTTTCCTCG 1474
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1664 TGCTTTTTAACCCACTCCCTGGATGCATTTTCCCTCCTTGCATTTCCCTCTTTTCCTG 1723

Qy   1475 AGTTCATCTAGAGAATCTGCACTATGTTTTTCCCTTTTGTCTTGAGATGAAAGTTTTA 1534
```

```
Db   1964 TTGACTTGCCTTTTGTTCTTTCCTTCATCTTAGTCCCTCTTAGTCATGCAGTATGGTCATT 2023
Qy   1775 GCTAGGTAGAGGGTATGTCCTTTATGTAAATGGCCACCGCATTTAGTAGTATTACATAAACTTT 1834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2024 GCTAGGTAGAGGGTATGTCCTTTATGTAAATGGCCACCGCATTTAGTAGTATTACATAAACTTT 2083

Qy   1835 CTTTTAACAATCTGTGCATAGTACATGCTGCTCTGTTCCATTTAGAGATTTGACAGAGGT 1894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2084 CTTTTAACAATCTGTGCATAGTACATGCTGCTCTGTTCCATTTAGAGATTTGACAGAGGT 2143

Qy   1895 TTCAGTTTAGTATACTCAAATCTTATTTTAGTGCTTGGGAAATCAAATCAGAATATCACA 1954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2144 TTCAGTTTAGTATACTCAAATCTTATTTTAGTGCTTGGGAAATCAAATCAGAATATCACA 2203

Qy   1955 TCCTCTCCAATTCTCTCTTACTCCAAATTGCTGGGAAACTTCATGTTACTAACTTGTTG 2014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2204 TCCTCTCCAATTCTCTCTTACTCCAAATTGCTGGGAAACTTCATGTTACTAACTTGTTG 2263

Qy   2015 CTCTAACTCTGCCATCTTGGTTTCCCCATCCCTCTCTCCTCCATGGTACGTGTGCTCCT 2074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2264 CTCTAACTCTGCCATCTTGGTTTCCCCATCCCTCTCTCCTCCATGGTACGTGTGCTCCT 2323

Qy   2075 AATATTAGCGTTGGTTGAGATTTCAGTGGTGTCCAATAATTCCTCTTCCCTCTGGTTGCCTT 2134
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2324 AATATTAGCGTTGGTTGAGATTTCAGTGGTGTCCAATAATTCCTCTTCCCTCTGGTTGCCTT 2383

Qy   2135 TCCTGAGATAATCCACTAAGAAGAATATTTTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGA 2194
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2384 TCCTGAGATAATCCACTAAGAAGAATATTTTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGA 2443

Qy   2195 AATTATCAACTGTGCACAAGGAAAAAAATAGATATGTGAAAGGTTCACGTAAATTTCCTC 2254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2444 AATTATCAACTGTGCACAAGGAAAAAAATAGATATGTGAAAGGTTCACGTAAATTTCCTC 2503

Qy   2255 ACATCACAGAGATAAAATTCAGAAGGAGAAAACACAGACCAAAGAGAAGTATCTAAG 2314
          ||||||||||||  |||||||||||||||||||||||||||||||||||||||||||
Db   2504 ACATCACAGAGATTAAAATTCAGAAGGAGAAAACACAGACCAAAGAGAAGTATCTAAG 2563

Qy   2315 ACCAAAGGGATGTGTGTTTTATTAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTAC 2374
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2564 ACCAAAGGGATGTGTGTTTTATTAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTAC 2623

Qy   2375 TTGTAAATAACTAGAAATAACATGATTTAGTCATAATTGTGAAAAATAATAATAATTTTT 2434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2624 TTGTAAATAACTAGAAATAACATGATTTAGTCATAATTGTGAAAAATAATAATAATTTTT 2683

Qy   2435 CTTGGATTTATGTTCTGTATCGTGTGAAAAAATAAAATTCTTATAAAACTC 2484
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2684 CTTGGATTTATGTTCTGTATCGTGTGAAAAAATAAAATTCTTATAAAACTC 2733
```

RESULT 3
ADE25678
ID   ADE25678 standard; cDNA; 2741 BP.
XX
AC   ADE25678;
XX
DT   29-JAN-2004 (first entry)
XX
DE   Human cDNA differentially expressed in foam cells #82.
XX
KW   Human; ss; differential expression; foam cell; LPS; lipopolysaccharide;
KW   cardiovascular disease; atherosclerosis.
XX
OS   Homo sapiens.
XX
PN   US2003194721-A1.
XX
PD   16-OCT-2003.
XX
PF   18-SEP-2002; 2002US-00247671.
XX
PR   19-SEP-2001; 2001US-0323784P.

```
Db        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1724 AGTTCATACTAGAGAATCTGCACTATGTTTTTCCCTTTTGTCTTGAGATGAAAGTTTA 1783
Qy   1535 AAATAATCCACCTCTGTCATTCCACTTCCACTCTCTGAACATCCCAAGCTGTATCCCTGGCCTCT 1594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1784 AAATAATCCACCTCTGTCATTCCACTTCCACTCTCTGAACATCCCAAGCTGTATCCCTGGCCTCT 1843
Qy   1595 TTTCTCAGACTATGTTTCTTTACTTGGGACCTAGAACTGGATTGGCATTGCTCCCT 1654
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1844 TTTCTCAGACTATGTTTCTTTACTTGGGACCTAGAACTGGATTGGCATTGCTCCCT 1903
Qy   1655 GATCAGATGAGACCTTTGATTATTTGCCCCCTCCTTAGGACCTTACACTCCGTCTTTCT 1714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1904 GATCAGATGAGACCTTTGATTATTTGCCCCCTCCTTAGGACCTTACACTCCCGTCTTTCT 1963
Qy   1715 TTGACTTGCCTTTTGTTTCTTTCCTTCATCTTAGTCCCTCTTCATGCAGTATGGGTCATT 1774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1964 TTGACTTGCCTTTTGTTTCTTTCCTTCATCTTAGTCCCTCTTCATGCAGTATGGGTCATT 2023
Qy   1775 GCTAGGTAGAGGTATGTCCTTTTATGTAATGGCCACCGCATTAGTATTACATAAACTTT 1834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2024 GCTAGGTAGAGGTATGTCCTTTTATGTAATGGCCACCGCATTAGTATTACATAAACTTT 2083
Qy   1835 CTTTAACAATCTGTGCATAGTACATGCTGCTCTGTTCCATTGACGATTGACAGAGGT 1894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2084 CTTTAACAATCTGTGCATAGTACATGCTGCTCTGTTCCATTGACGATTGACAGAGGT 2143
Qy   1895 TTCAGTTTAGTATACTCAAATCTTATTTAGTGCTTGGGAAATCAATTCAGAATATCACA 1954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2144 TTCAGTTTAGTATACTCAAATCTTATTTAGTGCTTGGGAAATCAATTCAGAATATCACA 2203
Qy   1955 TCCTCTCCAATTCTCTCTACTTCAAATGCTGGGAAACTCTCCATGTACTAACTTGTGTG 2014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2204 TCCTCTCCAATTCTCTCTCTACTTCAAATGCTGGGAAACTCTCCATGTTACTAACTTTGTTG 2263
Qy   2015 CTCTAACTCTGCCATCTGGTTTCCCCATCCCTTCCTCCTCCATGTACGTGTGCTCCT 2074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2264 CTCTAACTCTGCCATCTGGTTTCCCCATCCCTTCCTCCTCCATGTACGTGTGCTCCT 2323
Qy   2075 AATATTAGCGTTGGTTGGTGAGATTTCAGTGGTCCAATATTCCTCTCCCCTCGGTTGCCTT 2134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2324 AATATTAGCGTTGGTTGGTGAGATTTCAGTGGTCCAATATTCCTCTCCCCTCGGTTGCCTT 2383
Qy   2135 TCCTGAGATAATCCACTAAGAACATATTTTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGA 2194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2384 TCCTGAGATAATCCACTAAGAACATATTTTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGA 2443
Qy   2195 AATTATCAACTGCGCACAAGAGAAAAAATAGATATGTGAAAGTTCACGTAAATTCCTC 2254
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2444 AATTATCAACTGCGCACAAGAGAAAAAATAGATATGTGAAAGTTCACGTAAATTCCTC 2503
Qy   2255 ACATCACAGAAGATTAAAATTCAGAAAGGAGAAAAACAGACCAAGAGAAGTATCTAAG 2314
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2504 ACATCACAGAAGATTAAAATTCAGAAAGGAGAAAAACAGACCAAGAGAAGTATCTAAG 2563
Qy   2315 ACCAAAGGGATGTGTTTTATTATGTCCTAGGATGAAGAAATGCATAGAACATTGTAGTAC 2374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2564 ACCAAAGGGATGTGTTTTATTATGTCCTAGGATGAAGAAATGCATAGAACATTGTAGTAC 2623
Qy   2375 TTGTAAATAACTAGAAAATAACATGATTTAGTCATAATTGTGAAAAATAATAATAATTTTT 2434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2624 TTGTAAATAACTAGAAAATAACATGATTTAGTCATAATTGTGAAAAATAATAATAATTTTT 2683
Qy   2435 CTTGGATTTATGTTCTGTATCTGTGAAAAATAAATTCTTATAAAACTC 2484
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   2684 CTTGGATTTATGTTCTGTATCTGTGAAAAATAAATTCTTATAAAACTC 2733

RESULT 4
AAS60823
ID   AAS60823 standard; cDNA; 2527 BP.
XX
AC   AAS60823;
```

```
CC   The invention relates to 1046 novel nucleic acids which are used as
CC   markers for determining the sensitivity of a cancer cell to the
CC   anticancer agent TAXOL. Cancer cells can be treated with TAXOL when they
CC   are shown to express one of the 242 sensitivity markers or the cells are
CC   shown not to express one of the 804 resistance markers. The methods can
CC   be used to determine the effectiveness of TAXOL in the treatment of
CC   cancer cell growth in an individual. The markers can be used as targets
CC   in developing anti-cancer agents such as chemotherapeutic compounds. The
CC   markers can also be used as targets in developing treatments for cancer,
CC   particularly those cancers which display resistance to agents and exhibit
CC   expression of the markers. The anticancer agents developed by the novel
CC   method can be used to treat cancer. Probes based on the markers can be
CC   used to detect transcripts or genomic sequences corresponding to the
CC   markers, in the identification of cells or tissues which mis-express the
CC   protein. Cancers which may be targeted include carcinoma (e.g. squamous
CC   cell carcinoma, sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic
CC   leukaemia), lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's
CC   disease and tumours (e.g. glioma). The present sequence is one of the
CC   1046 novel cancer cell markers
XX
SQ   Sequence 2527 BP; 742 A; 431 C; 513 G; 824 T; 0 U; 17 Other;

Query Match             49.7%;  Score 1235.2;  DB 4;  Length 2527;
Best Local Similarity   98.6%;  Pred. No. 2.6e-268;
Matches 1298;  Conservative   0;  Mismatches 14;  Indels   5;  Gaps   5;

Qy   1116 AGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACATCTCGGTGGTTAGCA 1175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1170 AGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACATCTCGGTGGTTAGCA 1229
Qy   1176 ATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 1235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1230 ATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 1289
Qy   1236 GACGAATGCTATGGCTGCAGATGACTTCTTAGGAGGATTTGCCTTTTGAGCAAATAGGTA 1295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 GACGAATGCTATGGCTGCAGATGACTTCTTAGGAGGATTTGCCTTTTGAGCAAATAGGTA 1349
Qy   1296 GATGGTTGGTGGTGGTGTGGAAGCTTGGAAGCGGTCAGGTAGTAGTGGCTACTTTCTGCTTGGA 1355
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1350 GATGGTTGGTGGTGGTGTGGAAGCTTGGAAGCGGTCAGGTAGTAGTGGCTACTTTCTGCTTGGA 1409
```

RESULT 5
ABX63519
ID   ABX63519 standard; cDNA; 2068 BP.
XX
AC   ABX63519;
XX
DT   26-FEB-2003 (first entry)
XX
DE   Human cDNA #519 differentially expressed in activated vascular tissue.
XX
KW   Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;
KW   hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;
KW   gene therapy; vascular disease; cancer; coronary; artery disease;
KW   hypertension; diabetes; pre-eclampsia; restenosis;
KW   ischaemia-reperfusion injury; stroke.
XX
OS   Homo sapiens.
XX
PN   US2002137081-A1.
XX
PD   26-SEP-2002.
XX
PF   08-JAN-2002; 2002US-00044090.
XX
PR   28-JUL-2000; 2000US-0222469P.
PR   08-JAN-2001; 2001US-0260483P.
XX
PA   (BAND/) BANDMAN O.
XX
PI   Bandman O;
XX
DR   WPI; 2003-110597/10.
XX
PT   Combination for diagnosing, staging, treating, or monitoring the
PT   progression of treatment of a vascular disease, e.g. atherosclerosis,
PT   comprises several cDNAs that are differentially expressed in activated
PT   vascular tissue.
XX
PS   Claim 1; Page; 18pp; English.
XX
CC   This invention relates to a combination comprising several cDNAs that are
CC   differentially expressed in activated vascular tissue. The invention also
CC   discloses a high throughput method for detecting differentially expressed
CC   cDNAs in a sample. The cDNAs of the invention may have
CC   antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC   gynaecological; vasotropic and cerebroprotective activities and may be
CC   used in gene therapy. The cDNAs of the invention may be used in a high-
CC   throughput methods for detecting differential expression of one or more
CC   cDNAs in a sample, or screening several molecules or compounds to
CC   identify a molecule or compound that specifically binds a cDNA of the
CC   invention. A protein encoded by the cDNA may be used to screen several
CC   molecules or compounds to identify a ligand that specifically binds to
CC   the protein, or to produce or purify an antibody to the protein that can
CC   be used to detect a protein in a sample or purify a natural or
CC   recombinant protein from a sample. The nucleotides may be useful for
CC   diagnosing, staging, treating, or monitoring the progression of treatment
CC   of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
CC   disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion
CC   injury, restenosis, or stroke. The cDNAs can also be used for large-scale
CC   genetic or gene expression analysis of several new nucleic acid
CC   molecules. Antibodies to the proteins encoded by the cDNAs are useful for
CC   diagnosing pre-pathologic disorders, and chronic or acute diseases
CC   associated with abnormalities in the expression, amount or distribution
CC   of the protein. The present sequence represents a cDNA of the invention
CC   that is differentially expressed in activated vascular tissue. Note: The
CC   sequence data for this patent did not form part of the specification, but
CC   was obtained in electronic format directly from USPTO at
CC   http.seqdata.uspto.gov/sequence.html?DocID=20020137081
XX
SQ   Sequence 2068 BP; 657 A; 338 C; 452 G; 621 T; 0 U; 0 Other;

Query Match   18.4%; Score 456.6; DB 7; Length 2068;
Best Local Similarity   95.2%; Pred. No. 7.8e-93;

```
Qy      1356  TCTATTAAATACCTGGCAGCTCTCTGTCTTTCTGGGTTGTTGCCCTGTGATTAGTTCT  1415
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1410  TCTATTAAATACCTGGCAGCTCTCTGTCTTTTTTGGGTTGTTGCCCTGTGATTAGTTCT  1469

Qy      1416  GCTTTTAACCACTCCCTGGATGCATTTTCCCTCCTCTTTCCTTGGA  1475
              |||||||||||||||||||||||||||||||||||||||||||||
Db      1470  GCTTTTTAACCCACTCCCTGGATGCATTTTCCCTCCTCTTTCCTTGGA  1529

Qy      1476  GTTCATACTAGAGAATCTGACATATGTTTTTGTCTTGAGATGAAAGTTTAA  1535
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1530  GTTCATACTAGAGAATCTGACATATGTTTTTGTCTTGAGATGAAAGTTTAA  1589

Qy      1536  AATAATCCACCTCTGTCATTTCCACTCTCTGAACATCCCAAGCTGTATCCCTGGCCTCTT  1595
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1590  AATAATCCACCTCTGTCATTTCCACTCTCTGAACATCCCAAGCTGTATCCCTGGCCTCTT  1649

Qy      1596  TTCTCAGACTATGTTTCTTTACTTGGGACCTAGAACTGGATTGGCATTGCTCCTG  1655
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1650  TTCTCAGACTATGTTTCTTTACTTGGGACCTAGAACTGGATTGGCATTGCTCCTG  1709

Qy      1656  ATCAGATGAGACCTTGATTATTTGCCCCTTCCTTAGGACCTTACACTCCTGTCTTCTT  1715
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1710  ATCAGATGAGACCTTGATTATTTGCCCCTTCCTTAGGACCTTACACTCCTGTCTTCTT  1769

Qy      1716  TGACTTGCCTTTTGTTCTTCCTTCATCTTAGTCCCCTTCATGCAGTATGGTCATTG  1775
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1770  TGACTTGCCTTTTGTTCTTCCTTCATCTTAGTCCCCTTCATGCAGTATGGTCATTG  1829

Qy      1776  CTAGGTAGAGGTATGTCCTTTATGTAATGGCCACCGCATTAGTATTACATAAACTTTC  1835
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1830  CTAGGTAGAGGTATGTCCTTTATGTAATGGCCACCGCATTAGTATTACATAAACTTTC  1889

Qy      1836  TTTTAACAATCTGTGCATAGTACATGCTGCTCTCTGTTCATTTAGAGATTTGACAGAGGTT  1895
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1890  TTTTAACAATCTGTGCATAGTACATGCTGCTCTCTGTTCATTTAGAGATTTGACAGAGGTT  1949

Qy      1896  TCAGTTAGTATACTCAAATCTTATTTTAGTGCTTGGGAAATCAATTCAGAATATCACAT  1955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1950  TCAGTTAGTATACTCAAATCTTATTTTAGTGCTTGGGAAATCAATTCAGAATATCACAT  2009

Qy      1956  CCTCCCAATTCTCTCTTACTCAAATTGCTGGGAAACTTCATGTTACTAACTTTGTTGC  2015
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2010  CCTCTCAATTCTCTCTTACTCTTAGCTGGGAAACTTCATGTTACTAACTTTGTTGC  2069

Qy      2016  TCTAACTCTGCCATCTGGTTCCCATCCCTTCCTCCTCATGGTACGTGTGCTCCCA  2075
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2070  TCTAACTCTGCCATCTGGTTCCCCATCCCTTCCTCCTCATGGTACGTGTGCTCCTA  2129

Qy      2076  ATATTAGGCGTTGGTTGAGATTTCAGTGGTCCAATATTCCTTCCCCTCGGTTGCCTTT  2135
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2130  ATATTAGCGGTTGGTTGAGATTTCAGTGGTCCAATATTCCTTCCCCTCGGTTGCCTTT  2189

Qy      2136  CCTGAGATAATCCACTAAGAATATTTTGTGTTTCTTTTTCTCAGGGAATCTAAGGGAGAA  2195
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2190  CCTGAGATAATCCACTAAGAATATTTTGTGTTCTTTTTCTCAGGGAATCTAAGGGAGAA  2249

Qy      2196  ATTATCAACTGTGCACAAGGAAAAAAATAGATATGTGAAAGGTTCACGTAAATTTCCTCA  2255
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2250  ATTATCAACTGTGCACAAGGAAAAAAATAGATATGTGAAAGGTTCACGTAAATTTCCTCA  2309

Qy      2256  CATCACAGAAGAGATTAAAATTCAGAAAGGAGAAAACACAGACCAA-AGAGAAGTATCTAAG  2314
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      2310  CATCACAGAAGAGATTAAAATTCAGAAAGGAGAAAACACAGACCAAGAGAGAAGTATCTAAG  2369

Qy      2315  A-CCAAAGGGATGTGTTTATTAAT-GTCTAGGATGATGAAGAAATGCATA-GAACATTGTA-  2370
              | |||||||||||||||| ||||| |||||||||||||||||||||||| |||||||||
Db      2370  ACCCAAAGGGATGTGTTTTATTAATGGTCTAGGATGATGAAGAAATGCATAGGAACATTGTAG  2429

Qy      2371  GTACTTGTAAATAACTAGAAATAACATGATTTAGTCATAATTGTGAAAAATAATAAT  2427
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2430  GTACTTGTAAATAACTAGAAATAACACCTGGATTTGGGCCCATAATGTTGGGCCATTTGGGATTGTGAAAAAAANT  2486
```

Left column (RESULT 6):

```
Matches 471;    Conservative 0;    Mismatches 24;    Indels 0;    Gaps 0;

Qy    801 AAATATTTTTTGAATTAATAATCTGCTTATGTCTACCTTACAGTTTAATCCCATGGAAT 860
              ||||||| |||| |||||     ||||  |   |||||  ||||||||||||||||||
Db   1283 ACATATTCTATATCTTGAACGTAACATTCGTGATAGATACCAGTTTAATCCCATGGAAT 1342

Qy    861 CAATCAAATTAAATCATCATGACTACATTGAAGGACAGAATTCATT 920
              |||||||||||||||||||||||||||||||||||||||||||||
Db   1343 CAATCAAATTAAATCATCATGACTACATTGAAGGACAGAATTCATT 1402

Qy    921 GTGTGGCATTTGTATTTGATGCCAGCTCTATTCAATACTTCTCCTCTCAGATGATAGTAA 980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1403 GTGTGGCATTTGTATTTGATGCCAGCTCTATTCAATACTTCTCCTCTCAGATGATAGTAA 1462

Qy    981 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGGTACATGTGGCTTTGCTCA 1040
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1463 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGGTACATGTGGCTTTGCTCA 1522

Qy   1041 CTCATGTGGATAGCATGGATTTGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGGATGTG 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1523 CTCATGTGGATAGCATGGATTTGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGGATGTG 1582

Qy   1101 AGCCTGTGAGGGTCCAAGCTAGAGGAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACA 1160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1583 AGCCTGTGAGGGTCCAAGCTAGAGGAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACA 1642

Qy   1161 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA 1220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1643 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA 1702

Qy   1221 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1703 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1762

Qy   1281 TTGAGCAAATAGGTA 1295
              ||||||||||||| |
Db   1763 TTGAGCAAATAGGGA 1777
```

RESULT 6
AAX33941
ID   AAX33941 standard; DNA; 1335 BP.
XX
AC   AAX33941;
XX
DT   30-JUN-1999 (first entry)
XX
DE   Human HCMV inducible gene, SEQ ID NO 9.
XX
KW   HCMV inducible gene; cig; human; human cytomegalovirus; interferon;
KW   anti-viral therapy; anti-HCMV therapy; detection; diagnosis;
KW   drug screening; ds.
XX
OS   Homo sapiens.
XX
PN   WO9913075-A2.
XX
PD   18-MAR-1999.
XX
PF   08-SEP-1998; 98WO-US018638.
XX
PR   08-SEP-1997; 97US-0058180P.
PR   22-SEP-1997; 97US-0059725P.
XX
PA   (UYPR-) UNIV PRINCETON.
XX
PI   Zhu H, Cong J, Schenk T;
XX
DR   WPI; 1999-243729/20.
DR   P-PSDB; AAY05371.
XX
PT   New isolated human genes to obtain agents for antiviral therapy,
PT   particularly anti-HCMV therapy.

Right column:

XX
PS   Claim 2; Page 103-104; 184pp; English.
XX
CC   This sequence represents a human gene of the invention, that is induced
CC   to express by both HCMV and interferon (IFN), designated HCMV-inducible
CC   genes (cig or cigs). The invention also relates to genes that are
CC   repressed in the presence of HCMV infection, designated HCMV-repressible
CC   genes (crg or crgs). The products can be used to obtain agents which can
CC   be used for anti-viral therapy, particularly anti-HCMV therapy. They can
CC   also be used for the development of drugs that would allow for higher
CC   dosage IFN treatments without the concomitant toxicity normally
CC   associated with administering high levels of IFN. The products can also
CC   be used for detection, diagnosis and drug screening
XX
SQ   Sequence 1335 BP; 409 A; 232 C; 311 G; 383 T; 0 U; 0 Other;

```
Query Match     18.3%;   Score 455;  DB 2;  Length 1335;
Best Local Similarity   94.9%;   Pred. No. 1.6e-92;
Matches  470;   Conservative   0;   Mismatches  25;   Indels   0;   Gaps   0;

Qy    801 AAATATTTTTGAATTAATAATCTGCTTATGTCTTATGTCCACTTACAGTTTAATCCCATGGAAT 860
              |||||    |||||||||||||||||||||||||   ||| ||||||||||||||||||||
Db    797 ACATATTCTATATCTTGACACGGTAACATTCGTGATAGATACCAGTTTAATCCCATGGAAT 856

Qy    861 CAATCAAATTAAATCATCATGACTACATTGAAGGACAGAATTCATT 920
              ||||||||||||||||||||||||||||||||||||||||||||||
Db    857 CAATCAAATTAAATCATCATGACTACATTGAAGGACAGAATTCATT 916

Qy    921 GTGTGGCATTTGTATTTGATGCCAGCTCTATTCAATACTTCTCCTCTCAGATGATAGTAA 980
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    917 GTGTGGCATTTGTATTTGATGCCAGCTCTATTCAATACTTCTCCTCTCAGATGATAGTAA 976

Qy    981 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGGTACATGTGGCTTTGCTCA 1040
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    977 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGGTACATGTGGCTTTGCTCA 1036

Qy   1041 CTCATGTGGATAGCATGGATTTGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGGATGTG 1100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1037 CTCATGTGGATAGCATGGATTTGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGGATGTG 1096

Qy   1101 AGCCTGTGAGGGTCCAAGCTAGAGGAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACA 1160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1097 AGCCTGTGAGGGTCCAAGCTAGAGGAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACA 1156

Qy   1161 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA 1220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1157 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA 1216

Qy   1221 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1217 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1276

Qy   1281 TTGAGCAAATAGGTA 1295
              ||||||||||||| |
Db   1277 TTGAGCAAATAGGGA 1291
```

RESULT 7
ABN59995
ID   ABN59995 standard; cDNA; 1679 BP.
XX
AC   ABN59995;
XX
DT   28-JUN-2002 (first entry)
XX
DE   Novel human coding sequence SEQ ID NO: 406.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag; gene; ss.
XX
OS   Homo sapiens.

```
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US026015.
XX
PR   11-SEP-2000; 2000US-00659671.
XX
PA   (HYSB-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   P-PSDB; ABB97582.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis.
XX
PS   Claim 1; SEQ ID NO 406; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC   and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC   rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a coding sequence of the
CC   invention
XX
SQ   Sequence 1679 BP; 554 A; 275 C; 373 G; 477 T; 0 U; 0 Other;

Query Match             18.3%;  Score 455;  DB 6;  Length 1679;
Best Local Similarity   94.9%;  Pred. No. 1.7e-92;
Matches  470;           Conservative   0;  Mismatches  25;  Indels   0;  Gaps   0;

Qy    801 AAATATTTTTGAATTAATAATCTGCTTATGTCTACCTTACACAGTTTAATCCCATGGAAT 860
          ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    863 ACATATTCTATATCTTGAACGGTAACATTCGTGATAGATACCAGTTTAATCCCATGGAAT 922

Qy    861 CAATCAAATTAAATCATCATGATGACTACATTGATTCCCATCGCTGAAGGACAGAATTCATT 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    923 CAATCAAATTAAATCATCATGATGACTACATTGATTCCCATCGCTGAAGGACAGAATTCATT 982

Qy    921 GTGTGGCATTTGTATTTGATGGCCAGTCTATTCAATACTTCTCCCTCTCAGATGATGAGTAA 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    983 GTGTGGCATTTGTATTTGATGGCCAGTCTATTCAATACTTCTCCCTCTCAGATGATGAGTAA 1042

Qy    981 AGATCAAAAGAATTCAAAGGAGGAGTTGGTAAACGCTGGTGTGGTACATGTGGCTTTGCTCA 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1043 AGATCAAAAGAATTCAAAGGAGGAGTTGGTAAACGCTGGTGTGGTACATGTGGCTTTGCTCA 1102

Qy   1041 CTCATGTGGATAGCACATGGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1103 CTCATGTGGATAGCACATGGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1162

Qy   1101 AGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAAACTTGGATTTGCTCTTTCTGACA 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1163 AGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAAACTTGGATTTGCTCTTTCTGACA 1222

Qy   1161 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGAGCCTGTAAAGGATGTTCTAA 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1223 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGAGCCTGTAAAGGATGTTCTAA 1282

Qy   1221 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTCATGGGCTCGAGATGGACTTGGATTTGCCTT 1280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1283 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTCATGGGCTCGAGATGGACTTGGATTTGCCTT 1342

Qy   1281 TTGAGCAAATAGGTA 1295
```

```
Db   1343 TTGAGCAAATAGGGA 1357
          |||||||||||||| |
```

```
RESULT 8
AAN81259
ID   AAN81259 standard; DNA; 1332 BP.
XX
AC   AAN81259;
XX
DT   25-MAR-2003  (revised)
DT   24-OCT-1990  (first entry)
XX
DE   Sequence encoding antigenic protein specific to non-A non-B hepatitis.
XX
KW   Non-A non-B hepatitis; ss.
XX
OS   Synthetic.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..1332
FT                  /*tag= a
XX
PN   EP293274-A.
XX
PD   30-NOV-1988.
XX
PF   31-MAR-1988; 88EP-00400790.
XX
PR   31-MAR-1987; 87JP-00078313.
PR   04-JUN-1987; 87JP-00140586.
PR   10-NOV-1987; 87JP-00283990.
XX
PA   (MITU ) MITSUBISHI CHEM IND LTD.
XX
PI   Kamizono M,  Matsui R,  Teranishi Y,  Nakaanishi S,  Kitamura N;
XX
DR   WPI; 1988-339815/48.
DR   P-PSDB; AAP80669.
XX
PT   Non-A non-B hepatitis-specific antigenic protein - obtd. by recombinant
PT   DNA techniques from liver infected with non-A non-B hepatitis.
XX
PS   Disclosure; Page ?; 33pp; English.
XX
CC   The antigenic protein encoded by this sequence can be produced with low
CC   cost on a large scale using recombinant DNA methods. It is used for in
CC   vitro diagnosis of non-A non-B hepatitis. This DNA sequence can be used
CC   as a probe in hybridisation assays, detecting an infecton by the virus.
CC   See also AAN81260. (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC   on 25-MAR-2003 to correct PA field.) (Updated on 25-MAR-2003 to correct
CC   PI field.)
XX
SQ   Sequence 1332 BP; 408 A; 230 C; 314 G; 380 T; 0 U; 0 Other;

Query Match             18.1%;  Score 448.6;  DB 1;  Length 1332;
Best Local Similarity   94.1%;  Pred. No. 4.4e-91;
Matches  466;           Conservative   0;  Mismatches  29;  Indels   0;  Gaps   0;

Qy    801 AAATATTTTTGAATTAATAATCTGCTTATGTCTACCTTACAGTTTAATCCCATGGAAT 860
          ||||||||  |  | ||  ||||||||||||||  ||||||||||||||||||||||||
Db    797 ACATATCCTACATCTTGAACGGTAACATTCGTGATAGATACCAGTTTAATCCCATGGAAT 856

Qy    861 CAATCAAATTAAATCATCATGATGACTACATTGATTCCCATCGCTGAAGGACAGAATTCATT 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    857 CAATCAAATTAAATCATCATGATGACTACATTGATTCCCATCGCTGAAGGACAGAATTCATT 916

Qy    921 GTGTGGCATTTGTATTTGATGGCCAGTCTATTCAATACTTCTCCCTCTCAGATGATGAGTAA 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    917 GTGTGGCATTTGTATTTGATGGCCAGTCTATTCAATACTTCTCCCTCTCAGATGATGAGTAA 976

Qy    981 AGATCAAAAGAATTCGAAGGAGGAGTTGGTAAACGCTGGTGTGTGTACTATGTGGCTTTGCTCA 1040
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  977 AGATCAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTCGTACATGTGGCTTTGCTCA 1036
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1041 CTTCATGTGGATAGCAGCATGGATTTGGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1037 CTTCATGTGGATAGCAGCATGGATTTGGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1096
Qy 1101 AGCTGTGAGGTCCAAGTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTGACA 1160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1097 TGCCTGTGAGGTCCAAGTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTGACA 1156
Qy 1161 TCTCGGGTGGTTAGCAATATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA 1220
        ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db 1157 TCTCGGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA 1216
Qy 1221 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1280
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1217 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1276
Qy 1281 TTGAGCAAATAGGTA 1295
        ||||||||||| ||
Db 1277 TTGAGCAAATAGGGA 1291
```

```
CC  protein. Cancers which may be targeted include carcinoma (e.g. squamous
CC  cell carcinoma), sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic
CC  leukaemia), lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's
CC  disease and tumours (e.g. glioma). The present sequence is one of the
CC  1046 novel cancer cell markers
XX
SQ  Sequence 567 BP; 128 A; 102 C; 116 G; 220 T; 0 U; 1 Other;

Query Match          16.7%;    Score 415.4;  DB 4;  Length 567;
Best Local Similarity 98.4%;   Pred. No. 1e-83;
Matches 419;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

Qy 1116 AGCTAGGAGGAAGTCCAAAGAAAACTGGATTTGCTCTTCTGACATCTGGTGGTGGTTAGCA 1175
        ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db  142 AGCTAGGAGGAAGTCCAAAGAAAACTGGATTTGCTCTTCTGACATCTCGGTGGTGGTTAGCA 201
Qy 1176 ATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 1235
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  202 ATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 261
Qy 1236 GACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTTTTGAGCAAATAGGTA 1295
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  262 GACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTTTTGAGCAAATAGGTA 321
Qy 1296 GATGGTTGGTTGGTGTGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGGA 1355
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  322 GATGGTTGGTTGGTGTGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGGA 381
Qy 1356 TCTATTAAATACCTGGCAGCTCTCTGTCTTTTGTGGGTTGTGCCCTGTGTGATTAGTTCT 1415
        |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db  382 TCTATTAAATACCTGGCAGCTCTCTGTCTTTNTGTGGGTTGTGCCCTGTGTGATTAGTTCT 441
Qy 1416 GCTTTTTAACCCACCTCCCTGGATGCATTTTCCCCTCCTTGCATTCCCTCTTTTCCTGGA 1475
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  442 GCTTTTTAACCCACCTCCCTGGATGCATTTTCCCCTCCTTGCATTCCCTCTTTTCCTGGA 501
Qy 1476 GTTCATACTAGGAATCTGCACTATGTTTTCTTGCTTGGATGAAAGTTTTAA 1535
        |||||||||||||||||||||||||||||||||||||||||| |||||||
Db  502 AGTCATACTAGGAATCTGCACTATGTTTTCTTGCTTGGATGAAGATGAAAGTTTAA 561
Qy 1536 AATAAT 1541
        ||||||
Db  562 AATAAT 567
```

```
RESULT 9
AAS60407
ID  .AAS60407 standard; cDNA; 567 BP.
XX
AC  AAS60407;
XX
DT  29-JAN-2002 (first entry)
XX
DE  Human cancer agent-sensitive marker #138.
XX
KW  Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
KW  squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
KW  lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
KW  Hodgkin's disease; glioma; ss.
XX
OS  Homo sapiens.
XX
PN  WO200179556-A2.
XX
PD  25-OCT-2001.
XX
PF  13-APR-2001; 2001WO-US012132.
XX
PR  14-APR-2000; 2000US-0197538P.
XX
PA  (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI  Lillie J, Brown JL, Bolt A, Van Huffel C;
XX
DR  WPI; 2001-602933/68.
XX
PT  Novel nucleic acid, used as a marker to determine the effectiveness of
PT  using TAXOL to treat cancer cell growth in individuals;
XX
PS  Claim 1; Page 211; 527pp; English.
XX
CC  The invention relates to 1046 novel nucleic acids which are used as
CC  markers for determining the sensitivity of a cancer cell to the
CC  anticancer agent TAXOL. Cancer cells can be treated with TAXOL when they
CC  are shown to express one of the 242 sensitivity markers or the cells are
CC  shown not to express one of the 804 resistance markers. The methods can
CC  be used to determine the effectiveness of TAXOL in the treatment of
CC  cancer cell growth in an individual. The markers can be used as targets
CC  in developing anti-cancer agents such as chemotherapeutic compounds. The
CC  markers can also be used as targets in developing treatments for cancer,
CC  particularly those cancers which display resistance to agents and exhibit
CC  expression of the markers. The anticancer agents developed by the novel
CC  method can be used to treat cancer. Probes based on the markers can be
CC  used to detect transcripts or genomic sequences corresponding to the
CC  markers, in the identification of cells or tissues which mis-express the
```

```
RESULT 10
ABL80943
ID  ABL80943 standard; cDNA; 482 BP.
XX
AC  ABL80943;
XX
DT  17-MAY-2002 (first entry)
XX
DE  Human ovarian cancer related cDNA clone SEQ ID NO:3921.
XX
KW  Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS  Homo sapiens.
XX
PN  WO200192581-A2.
XX
PD  06-DEC-2001.
XX
PF  29-MAY-2001; 2001WO-US017756.
XX
PR  26-MAY-2000; 2000US-0207484P.
XX
PA  (CORI-) CORIXA CORP.
XX
PI  Algate PA, Harlocker SL, Jones R;
XX
DR  WPI; 2002-122075/16.
XX
```

PT   Composition for therapy and diagnosis of ovarian cancer comprising
PT   polypeptide of a ovarian tumor polypeptide; polynucleotide encoding
PT   polypeptide encoded by a polynucleotide
PT   polypeptide, antibody specific to polypeptide or T cell expressing
PT   polypeptide.
XX
PS   Claim 1; SEQ ID NO 3921; 489pp; English.
XX
CC   The present invention describes a composition (I) comprising: carriers
CC   and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC   polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC   from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC   (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC   or antigen presenting cells that express (II). (I) has cytostatic
CC   activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC   detecting ovarian cancer in a patient's biological sample preferably
CC   serum or ovarian tissue. The method comprises contacting a biological
CC   sample from a patient with (IV), detecting the amount of polynucleotide
CC   hybridising to (IV) and comparing the amount to a predetermined cutoff
CC   value and thereby detecting ovarian cancer in the patient, where the
CC   amount of polynucleotide hybridising to (IV) is detected preferably by
CC   polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC   useful for stimulating and/or expanding T cells specific for an ovarian
CC   tumour protein comprising contacting T cells with (III) or (II). (III) is
CC   useful in design and preparation of ribozyme molecules for inhibiting
CC   expression of the tumour polypeptides and proteins in tumour cells; and
CC   to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC   library using well known techniques
XX
SQ   Sequence 482 BP; 104 A; 90 C; 103 G; 185 T; 0 U; 0 Other;

Query Match   14.8%; Score 368; DB 6; Length 482;
Best Local Similarity 100.0%; Pred. No. 4.7e-73;
Matches 368; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1116 AGCTAGGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACATCTCGGTGGTTAGCA 1175
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   115 AGCTAGGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACATCTCGGTGGTTAGCA 174

Qy   1176 ATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 1235
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   175 ATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 234

Qy   1236 GACGAATGCTATGGGCTGCAGATGACTTCTAGAGGATTGCCTTTTGAGCAAATAGGTA 1295
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   235 GACGAATGCTATGGGCTGCAGATGACTTCTAGAGGATTGCCTTTTGAGCAAATAGGTA 294

Qy   1296 GATGGTTGGTGCGTGCGGAAGCTTGGAAGCCGGTCAGTAGTTGGCTACTTTCTGCTTGGA 1355
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   295 GATGGTTGGTGCGTGCGGAAGCTTGGAAGCCGGTCAGTAGTTGGCTACTTTCTGCTTGGA 354

Qy   1356 TCTATTAAATAACCTGGCAGCTCTCTGCTTTTGTGGGTTGTTGCCCTGTGATTAGTTCT 1415
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   355 TCTATTAAATAACCTGGCAGCTCTCTGCTTTTGTGGGTTGTTGCCCTGTGATTAGTTCT 414

Qy   1416 GCTTTTTAACCCACTCCCTGGATGCATTTTCCCTCCTGCATTTTCCCTCCTTTCCTTGGA 1475
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   415 GCTTTTTAACCCACTCCCTGGATGCATTTTCCCTCCTGCATTTTCCCTCCTTTCCTTGGA 474

Qy   1476 GTTCATAC 1483
           ||||||||
Db   475 GTTCATAC 482

RESULT 11
ABX63520
ID   ABX63520 standard; cDNA; 2515 BP.
XX
AC   ABX63520;
XX
DT   26-FEB-2003 (first entry)
XX
DE   Human cDNA #520 differentially expressed in activated vascular tissue.
XX

KW   Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;
KW   hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;
KW   gene therapy; vascular disease; cancer; coronary; artery disease;
KW   hypertension; diabetes; pre-eclampsia; restenosis;
KW   ischaemia-reperfusion injury; stroke.
XX
OS   Homo sapiens.
XX
PN   US2002137081-A1.
XX
PD   26-SEP-2002.
XX
PF   08-JAN-2002; 2002US-00044090.
XX
PR   28-JUL-2000; 2000US-0222469P.
PR   08-JAN-2001; 2001US-0260483P.
XX
PA   (BAND/) BANDMAN O.
XX
PI   Bandman O.
XX
DR   WPI; 2003-110597/10.
XX
PT   Combination for diagnosing, staging, treating, or monitoring the
PT   progression of treatment of a vascular disease, e.g. atherosclerosis,
PT   comprises several cDNAs that are differentially expressed in activated
PT   vascular tissue.
XX
PS   Claim 1; Page; 18pp; English.
XX
CC   This invention relates to a combination comprising several cDNAs that are
CC   differentially expressed in activated vascular tissue. The invention also
CC   discloses a high throughput method for detecting differentially expressed
CC   cDNAs in a sample. The cDNAs of the invention may have
CC   antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC   gynaecological; vasotropic and cerebroprotective activities and may be
CC   used in gene therapy. The cDNAs of the invention may be used in a high-
CC   throughput methods for detecting differential expression of one or more
CC   cDNAs in a sample, or screening several molecules or compounds to
CC   identify a molecule or compound that specifically binds a cDNA of the
CC   invention. A protein encoded by the cDNA may be used to screen several
CC   molecules or compounds to identify a ligand that specifically binds to
CC   the protein, or to produce or purify an antibody to the protein that can
CC   be used to detect a protein in a sample or purify a natural or
CC   recombinant protein from a sample. The nucleotides may be useful for
CC   diagnosing, staging, treating, or monitoring the progression of treatment
CC   of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
CC   disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion
CC   injury, restenosis, or stroke. The cDNAs can also be used for large-scale
CC   genetic or gene expression analysis of several new nucleic acid
CC   molecules. Antibodies to the proteins encoded by the cDNAs are useful for
CC   diagnosing pre-pathologic disorders, and chronic or acute diseases
CC   associated with abnormalities in the expression, amount or distribution
CC   of the protein. The present sequence represents a cDNA of the invention
CC   that is differentially expressed in activated vascular tissue. Note: The
CC   sequence data for this patent did not form part of the specification, but
CC   was obtained in electronic format directly from USPTO at
CC   http.seqdata.uspto.gov/sequence.html?DocID=2002137081
XX
SQ   Sequence 2515 BP; 747 A; 430 C; 538 G; 800 T; 0 U; 0 Other;

Query Match   14.3%; Score 354; DB 7; Length 2515;
Best Local Similarity 99.2%; Pred. No. 1.1e-69;
Matches 387; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

Qy   1116 AGCTAGGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACATCTCGGTGGTTAGCA 1175
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2128 AGCTAGGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACATCTCGGTGGTTAGCA 2187

Qy   1176 ATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 1235
           ||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   2188 ATTATTCCTCTGAGTGGGAGCTGGAGCCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 2247

```
Qy   1236  GACGAATGCTATGGGCTGCAGATGACTTCTTAGAGATTGCCTTTTGAGCAAATAGGTA  1295
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2248  GACGAATGCTATGGGCTGCAGATGACTTCTTAGAGATTGCCTTTTGAGCAAATAGGTA  2307

Qy   1296  GATG-GTTTGTGGTGTGGAAGCTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGG  1354
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2308  GATGTGTTTGGTGGTGTGGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGG  2367

Qy   1355  ATCTATTAAATACCTGCCAGCTCGTCTTTTGTGGGTGTGCCCTGATTAGTTC  1414
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2368  ATCTATTAAATACCTGCCAGCTCTCTGTCTTTTGTGGGTGTGCCCTGATTAGTTC  2427

Qy   1415  TGCTTTTTAACCACTCCCTGGATGCATTTTCCCTCCTTGCATTTCCCTCTTTTCCTGG  1474
           ||||||||||||||||||||| ||||||||||||||||||||  ||||||||||||||
Db   2428  TGCTTTTTAACCCACTCCCTGGATGCA-TTTTCCCTCCTTGCA-TTCCCTCTTTTCCTGG  2485

Qy   1475  AGTTCATCATACTAGAGAATCTGCACTATGTTT  1504
           |||||||||||||||||||||||||||||||
Db   2486  AGTTCATCATACTAGAGAATCTGCACTATGTTT  2515


RESULT 12
AAS60100
ID   AAS60100 standard; cDNA; 580 BP.
XX
AC   AAS60100;
XX
DT   29-JAN-2002  (first entry)
XX
DE   Human cancer agent-sensitive marker #101.
XX
KW   Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
KW   squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
KW   lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
KW   Hodgkin's disease; glioma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200179556-A2.
XX
PD   25-OCT-2001.
XX
PF   13-APR-2001; 2001WO-US012132.
XX
PR   14-APR-2000; 2000US-0197538P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J, Brown JL, Bolt A,  Van Huffel C;
XX
DR   WPI; 2001-602933/68.
XX
PT   Novel nucleic acid, used as a marker to determine the effectiveness of
PT   using TAXOL to treat cancer cell growth in individuals.
XX
PS   Claim 1; Page 119; 527pp; English.
XX
CC   The invention relates to 1046 novel nucleic acids which are used as
CC   markers for determining the sensitivity of a cancer cell to the
CC   anticancer agent TAXOL. Cancer cells can be treated with TAXOL when they
CC   are shown to express one of the 242 sensitivity markers or the cells are
CC   shown not to express one of the 804 resistance markers. The methods can
CC   be used to determine the effectiveness of TAXOL in the treatment of
CC   cancer cell growth in an individual. The markers can be used as targets
CC   in developing anti-cancer agents such as chemotherapeutic compounds. The
CC   markers can also be used as targets in developing treatments for cancer,
CC   particularly those cancers which display resistance to agents and exhibit
CC   expression of the markers. The anticancer agents developed by the novel
CC   method can be used to treat cancer. Probes based on the markers can be
CC   used to detect transcripts or genomic sequences corresponding to the
CC   markers, in the identification of cells or tissues which mis-express the
CC   protein. Cancers which may be targeted include carcinoma (e.g. squamous
CC   cell carcinoma), sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic
```

```
CC   leukaemia), lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's
CC   disease and tumours (e.g. glioma). The present sequence is one of the
CC   1046 novel cancer cell markers
XX
SQ   Sequence 580 BP; 125 A; 111 C; 130 G; 212 T; 0 U; 2 Other;

Query Match              13.3%;  Score 331.2;  DB 4;  Length 580;
Best Local Similarity    96.4%;  Pred. No. 9.8e-65;
Matches 380;  Conservative  0;  Mismatches 10;  Indels  4;  Gaps  4;

Qy   1123  GGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACATCTCGGTGGTTAGCAATTATTC  1182
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    188  GGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACATCTCGGTGGTTAGCAATTATTC   247

Qy   1183  CTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCCTGAGACGAAT  1242
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    248  CTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCCTGAGACGAAT   307

Qy   1243  GCTATGGGCTGCAGATGACTTCTTAGAGATTGCCTTTTGAGCAAATAGGTAGTAGATGGTT  1302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    308  GCTATGGGCTGCAGATGACTTCTTAGAGATTGCCTTTTGAGCAAATAGGTAGTAGATGGTT   367

Qy   1303  TGGTGGTGT-GGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGGATCTATT  1361
           |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    368  TGGTGTGTGGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGGATCTATT   427

Qy   1362  AAATACCTGCCAGCTCTCTGTCTTTTGTGGGTGTGCCCTGATTAGTTCTGCTTTT  1421
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    428  AAATACCTGGCAGCTCTCTGTC-TTTTGTGGGTGTGCCCTGAATAGTTCTGCTTTT   486

Qy   1422  TAACCCACTCCCTGGATGCATTTTCCCTCCTTGCATTTCCCTCTTTTCCTGG-AGTTCA  1480
           ||||||||||||||||||||||||||||||||| ||||||||||||||||||| |||||
Db    487  TAACCCACTCCCTGGATGCATTTTTNCCTTCTTGCATTTCCCTCTTTTCCTGGAAGTTCA   546

Qy   1481  TACT-AGAGAATCTGCACTATGTTTTTTCCCTTTT  1513
           |||| |||||||||||||||||||||| ||| |||
Db    547  TACTAANAGAATCTGCACTAATGGTTTTCCCTTT   580


RESULT 13
AAH34468
ID   AAH34468 standard; cDNA; 977 BP.
XX
AC   AAH34468;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:1550.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; chromosome 1; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US026524.
XX
PR   29-SEP-1999;  99US-0157137P.
PR   03-NOV-1999;  99US-0163280P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   P-PSDB; AAG75063.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers.
XX
```

```
PS   Claim 1; Page 3193-3194; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC   proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene therapy
CC   and vaccine production. N and P may be used in the prevention, diagnosis
CC   and treatment of diseases associated with inappropriate P expression. For
CC   example, N and P may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of P by expressing inactive proteins or to
CC   supplement the patients own production of P. Additionally, N may be used
CC   to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC   into a host cell and culturing the cell to express the proteins. N and P
CC   can be used in the prevention, diagnosis and treatment of colorectal
CC   carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC   sequences used in the exemplification of the present invention. N.B.
CC   Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC   time of publication, meaning no sequences are present for SEQ ID NO:1027
CC   to 1052, 7921 and 7922
XX
SQ   Sequence 977 BP; 353 A; 158 C; 202 G; 259 T; 0 U; 5 Other;

Query Match        12.7%;   Score 315.4;   DB 4;   Length 977;
Best Local Similarity  82.1%;   Pred. No. 4.2e-61;
Matches  372;  Conservative   1;  Mismatches   76;  Indels   4;  Gaps   1;

Qy      843 AGTTAATCCCATGGAATCAATCAAATTAAATCATCATGACTACATTGATTCCCATCGC 902
                |||| || | |   |||| ||   ||| | |||||  |  ||| | || || ||
Db      202 AGATTCCACCTAGAAANGGGGAAAGCTNTTCCNGGGAACACCATTTATACCCCCACAC 261

Qy      903 TGAAGGACAGAATTCATTGTGTGCCATTTGTATTTGATGCCAGTCTATTCAATACTTCT 962
                ||||||  | | | ||   || ||  ||||   ||
Db      262 AAAATAAATAGCATGAGCTGTGT---TTTAGAGGAGATAGGGTGCCAAACCAAATTCACT 317

Qy      963 CCTCTCAGATGATAGTAAGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGG 1022
                |||||    |||  || || |  || |||  |||   | || || ||| ||||||||
Db      318 CCTCTCAGATGATAGTAAAGATCAAAAGNATTCGAAGGGAGTTGGTAAACGCTGGTGTGG 377

Qy     1023 TACATGTGGCTTTGCTCACTCATGTGGATAGCATGGATTGATTACAAAAGGTGACCTTA 1082
                ||||||| ||||  || |||| ||||| |||| |||| |||| |||||||| || ||
Db      378 TACATGTGGCTTTSCTCACTCATGTGGATAGCATGGATTTGATTACAAAAGGTGACCTA 437

Qy     1083 TAGAAATAGAGAGATGTGAGCCTGTGAGGTCCAAGCTAGAGGAGTCCAAAGAAAACTTG 1142
                ||||||||||||||| ||| ||||| ||| || |||||| | || ||||||| | ||||
Db      438 TAGAAATAGAGAGATGTGAGCCTGTGAGGTCCAAGCTAGAGGAGTCCAAAGAAAACTTG 497

Qy     1143 GATTGCTCTTTCTGACATCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACC 1202
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      498 GATTGCTCTTTCTGACATCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACC 557

Qy     1203 CTGTAAAGGATGTTCTAATTCTTTCTGCTCTGAGACGAATGCTATGGGGTGCAGATGACT 1262
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      558 CTGTAAAGGATGTTCTAATTCTTTCTGCTCTGAGACGAATGCTATGGGGTGCAGATGACT 617

Qy     1263 TCTTAGAGGATTTGCCTTTTGAGCAAATAGGTA 1295
                |||||||||||||||||||||||||||||||||
Db      618 TCTTAGAGGATTTGCCTTTTGAGCAAATAGGGA 650


RESULT 14
AAS27231
ID   AAS27231 standard; cDNA; 3276 BP.
XX
AC   AAS27231;
XX
DT   07-NOV-2001 (first entry)
XX
DE   cDNA encoding novel signal transduction pathway protein, Seq ID 266.
XX
KW   Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW   antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW   immune system disorder; rheumatoid arthritis; inflammatory condition;
```

```
KW   organ transplant rejection; infection; hepatitis C; blood disorder;
KW   sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW   cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW   reproductive system; gastrointestinal; liver disorder; AIDS; ss;
XX
OS   Homo sapiens.
XX
FH   WO20015733-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001312.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184464P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214486P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0209964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231411P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
```

PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0239937P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241766P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.

PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0256678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465460/50.
DR P-PSDB; AAU17314.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
XX disorders and neuronal disorders.
PS Claim 1; SEQ ID NO 266; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II), (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative disorders
CC (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.
CC Alzheimer's disease, Parkinson's disease), chromosomal abnormalities
CC (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g.
CC glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),
CC respiratory disorders, dermatological disorders, in wound healing,
CC epithelial cell proliferation, endocrine disorders (e.g. Addison's
CC disease), reproductive system disorders, gastrointestinal disorder
CC (inflammatory disorders), liver disorders (cirrhosis), as stimulators of
CC B-cell responsiveness to pathogens, activators of T-cells, to induce
CC higher affinity antibodies, and as a means to induce tumour proliferation
CC in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAS26976-
CC AAS27850 represent novel signal transduction pathway protein coding
CC sequences and PCR primers of the invention
XX

Query Match   12.5%;   Score 311; DB 4; Length 3276;
Best Local Similarity   95.5%;   Pred. No. 6e-60;
Matches 320; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 2150 CTAAGAATATTTGTGTTCTTTTCTCAGGGAATCTAAGGGAGGAAATTATCAACTGTGC 2209
          |||||| |||| |||||| || |||||||||||||||||||||||||||||||
Db 2923 CTTAGAGGATTTGCCTTTGAGCAAATAGGGAATCTAAGGGAGGAAATTATCAACTGTGC 2982

Qy 2210 ACAAGGAAAAAAAATAGATATGTGAAAGGTTCACGTAAATTCCTCACATCACAGAAGATT 2269
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2983 ACAAGGAAAAAAAATAGATATGTGAAAGGTTCACGTAAATTCCTCACATCACAGAAGATT 3042

Qy 2270 AAAATTCAGAAGGAGAAAACACAGACCAAAGAAGAGTATCTAAGACCAAAGGGATGTGT 2329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3043 AAAATTCAGAAGGAGAAAACACAGACCAAAGAAGAGTATCTAAGACCAAAGGGATGTGT 3102

Qy 2330 TTTATTAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTACTTGTAAATAACTAGA 2389
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3103 TTTATTAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTACTTGTAAATAACTAGA 3162

Qy 2390 AATAACATGATTTAGTCATAATTGTGAAAAATAATAATATATTTTCTTGGATTTATGTTC 2449
          ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db 3163 AATAACATGATTTAGTCATAATTGTGAAAAATAATAATATATTTTTCTTGGATTTATGTTC 3222

Qy 2450 TGTATCTGTGAAAAAATAAATTCTTATAAAACTC 2484
          ||||||||||||||||||||||||||||||||||
Db 3223 TGTATCTGTGAAAAAATAAATTCTTATAAAACTC 3257

```
RESULT 15
ADB93409
ID   ADB93409 standard; cDNA; 3276 BP.
XX
AC   ADB93409;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human cDNA encoding a novel protein #256.
XX
KW   ss; gene; human; autoimmune disease; Parkinson's disease; silicosis;
KW   gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;
KW   immunosuppressive agent; adjuvant; enhance immune response;
KW   higher affinity antibody induction;
KW   increased serum immunoglobulin concentration.
XX
OS   Homo sapiens.
XX
PN   US2002168711-A1.
XX
PD   14-NOV-2002.
XX
PF   17-JAN-2001; 2001US-00764868.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216680P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-719985/68.
DR   P-PSDB; ADB94022.
XX
PT   New isolated polypeptide useful for diagnosing and treating
PT   immunosuppressive conditions such as autoimmune disease and Parkinson's
PT   disease.
XX
PS   Claim 3; SEQ ID NO 266; 345pp; English.
XX
CC   The invention relates to an isolated polypeptide. The polypeptide is
CC   useful for diagnosing a pathological condition or a susceptibility to a
CC   pathological condition in a subject, by determining the presence or
CC   amount of expression of the polypeptide in a biological sample and
CC   diagnosing a pathological condition or a susceptibility to a pathological
CC   condition based on the presence or amount of expression of the
CC   polypeptide. The polypeptide is also useful for identifying a binding
CC   partner to the polypeptide, which involves contacting the polypeptide
CC   with a binding partner and determining whether the binding partner
CC   effects an activity of the polypeptide. The polypeptide or the nucleic
CC   acid encoding the polypeptide is useful for preventing, treating, or
CC   ameliorating a medical condition, which involves administering the
CC   polypeptide or the nucleic acid to a mammalian subject. The nucleic acid
CC   is useful for diagnosing a pathological condition or a susceptibility to
CC   a pathological condition in a subject, which involves determining the
CC   presence or absence of a mutation in The nucleic acid, and diagnosing a
CC   pathological condition or susceptibility to a pathological condition
CC   based on the presence or absence of the mutation. The polypeptide, the
CC   nucleic acid and an antibody to the polypeptide are useful for treating
CC   autoimmune disease, Parkinson's disease, silicosis, gastrointestinal
CC   disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,
CC   the nucleic acid and the antibody are useful as immunosuppressive agents,
CC   as adjuvants to enhance immune responses, and as agents to induce higher
CC   affinity antibodies and increase serum immunoglobulin concentration. The
CC   present sequence represents cDNA encoding a novel human protein. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format direct from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20020168711.
XX
SQ   Sequence 3276 BP; 1075 A; 495 C; 669 G; 1037 T; 0 U; 0 Other;

Query Match             12.5%;  Score 311;  DB 9;  Length 3276;
Best Local Similarity   95.5%;  Pred. No. 6e-60;
Matches  320; Conservative   0; Mismatches  15; Indels   0; Gaps   0;

Qy       2150 CTAAGAATATTTGTTTCTTTTCTCAGGGAATCTAAGGGAGGAAATTATCAACTGTGC 2209
              |||||||||||| |||||||| |||||||||||||||||||||||||||||||||||
Db       2923 CTTAGAGGATTTGCCTTTGAGCAAATAGGGAATCTAAGGGAGGAAATTATCAACTGTGC 2982

Qy       2210 ACAAGGAAAAAAATAGATATGTGAAAGGTTCACGTAAATTCCTCACATCACGAAGATT 2269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2983 ACAAGGAAAAAAATAGATATGTGAAAGGTTCACGTAAATTCCTCACATCACGAAGATT 3042

Qy       2270 AAAATTCAGAAGGAGAAAACACAGACCAAAGAGAATCTAAGACAAAGGGATGT 2329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3043 AAAATTCAGAAGGAGAAAACACAGACCAAAGAGAATCTAAGACAAAGGGATGTGT 3102

Qy       2330 TTTATTAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTACTTGTAAATAACTAGA 2389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3103 TTTATTAATGTCTAGGATGAAGAAATGCATAGAACATTGTAGTACTTGTAAATAACTAGA 3162

Qy       2390 AATAACATGATTTAGTCATAATTGTGAAAAATATAATAATTTTCTTGGATTTATGTTC 2449
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3163 AATAACATGATTTAGTCATAATTGTGAAAAATATAATAATTTTCTTGGATTTATGTTC 3222

Qy       2450 TGTATCTGTGAAAAATAAATTCTTATAAAACTC 2484
```

us-10-051-835-18.rng

, Mon Apr 26 09:48:41 2004

Db   3223 TGTATCTGTGAAAAATAAATTTCTTATAAAACTC 3257

Search completed: April 25, 2004, 08:46:17
Job time : 665.01 secs

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 23:54:27 ; Search time 6540.9 Seconds
(without alignments)
16460.143 Million cell updates/sec

```
o    GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
```

Title: US-10-051-835-18
Perfect score: 2484
Sequence: 1 gtttattcctgtattataac........ataaattcttataaaactc 2484

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151699S residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   GenEmbl:*
```
1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pri:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1369	55.1	3062	9	HSM805959	BX537508 Homo sapi
2	1369	55.1	151265	9	AC104837	AC104837 Homo sapi
3	1235.2	49.7	2527	6	AX285019	AX285019 Sequence
4	456.6	18.4	1714	9	BC022870	BC022870 Homo sapi
5	455	18.3	1679	6	AX405991	AX405991 Sequence
6	448.6	18.1	1062	6	E01792	E01792 DNA sequenc
7	448.6	18.1	1062	6	E01977	E01977 DNA encodin
8	448.6	18.1	1332	6	E01791	E01791 DNA sequenc
9	448.6	18.1	1332	6	E01975	E01975 DNA encodin
10	448.6	18.1	1665	6	CHPP44	D90034 Pan troglod
11	448.6	18.1	1667	6	E01793	E01793 DNA sequenc
12	448.6	18.1	1667	6	E01978	E01978 DNA encodin
13	415.4	16.7	567	6	AX284603	AX284603 Sequence
14	364	14.7	483	9	D28908S8	D28915 Homo sapien
15	331.2	13.3	580	6	AX284296	AX284296 Sequence
16	283.2	11.4	2797	10	MMU299405	AJ299405 Mus muscu
17	272	11.0	272	6	AX378997	AX378997 Sequence
18	252	10.1	372	9	D28908S7	D28914 Homo sapien
19	240	9.7	240	6	AR240306	AR240306 Sequence
20	232.4	9.4	341	9	D28908S5	D28912 Homo sapien
21	223.8	9.0	1793	10	BC026901	BC026901 Mus muscu
22	196.8	7.9	2058	9	AB000115	AB000115 Homo sapi
23	196.8	7.9	2062	6	BD231210	BD231210 Human cyt
24	195.2	7.9	2058	9	BC015932	BC015932 Homo sapi
25	195.2	7.9	4757	9	HSM803928	AL832618 Homo sapi
26	165.6	6.7	2228	10	BC024930	BC024930 Mus muscu
27	160	6.4	187323	10	AC123055	AC123055 Mus muscu
28	158	6.4	2279	10	AF336221	AF336221 Mus muscu
29	143.6	5.8	252914	2	AC099264	AC099264 Rattus no
30	106.8	4.3	288	9	D28908S6	D28913 Homo sapien
31	101.2	4.1	489	6	AX395818	AX395818 Sequence 14
32	90.6	3.6	7218	2	I66494	I66494 Sequence 14
33	84	3.4	218015	2	AC108332	AC108332 Rattus no
34	72	2.9	112843	9	AP003165	AP003165 Homo sapi
35	71.6	2.9	105360	9	AC093218	AC093218 Homo sapi
36	71.6	2.9	198784	9	AC021318	AC021318 Homo sapi
37	69.2	2.8	76568	3	MBREV	AF538053 Monosiga
38	66	2.7	82291	9	AC005376	AC005376 Homo sapi
39	65.2	2.6	169318	9	AL592425	AL592425 Homo sapi
40	65.2	2.6	173660	9	AC063979	AC063979 Homo sapi
41	64.6	2.6	49573	9	AL583829	AL583829 Human DNA
42	64.6	2.6	93791	2	AC138073	AC138073 Homo sapi
43	64.2	2.6	145063	9	AC027069	AC027069 Homo sapi
44	64.2	2.6	154068	9	AL356292	AL356292 Human DNA
45	63.4	2.6	155260	9	AC074085	AC074085 Homo sapi

ALIGNMENTS

RESULT 1
HSM805959
```
LOCUS       HSM805959              3062 bp    mRNA    linear   PRI 17-JUN-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686L21223 (from clone DKFZp686L21223).
ACCESSION   BX537508
VERSION     BX537508.1  GI:31873567
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3062)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Direct Submission
```

Pred. No. is the number of results predicted by chance to have a

```
JOURNAL     Submitted (17-JUN-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp686L21223) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1..3062
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /map="1p31.1"
                     /clone="DKFZp686L21223"
                     /tissue_type="human small intestine"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_SfiI; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
     polyA_signal    3013..3018
     polyA_site      3039
ORIGIN

Query Match             55.1%;  Score 1369;  DB 9;  Length 3062;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-260;
  Matches 1369;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1116 AGCTAGAGGAAGTCCAAAGAAACCTGGATTGCTCTTCTTCTGACATCTCGGTGGTTAGCA 1175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1665 AGCTAGAGGAAGTCCAAAGAAACCTGGATTGCTCTTCTTCTGACATCTCGGTGGTTAGCA 1724

Qy   1176 ATTATTCCTCGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 1235
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1725 ATTATTCCTCGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGA 1784

Qy   1236 GACGAATGCTATGGGCTCGCAGATGACTTCTTAGAGGATTTGCCTTTGAGCAAATAGTA 1295
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1785 GACGAATGCTATGGGCTCGCAGATGACTTCTTAGAGGATTTGCCTTTGAGCAAATAGTA 1844

Qy   1296 GATGGTTTGGTGGTGTGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTCTGCTTGA 1355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1845 GATGGTTTGGTGGTGTGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTCTGCTTGA 1904

Qy   1356 TCTATTAAATACCTGGCAGCTCTCTGTCTTTTGTGGGTGTGTTGCCCTGTGATTAGTTCT 1415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1905 TCTATTAAATACCTGGCAGCTCTCTGTCTTTTGTGGGTGTGTTGCCCTGTGATTAGTTCT 1964

Qy   1416 GCTTTTTAACCCACTCCCTGGATGCATTTTCCCTCCTGCATTTCCCTCTTTTTCCTGA 1475
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1965 GCTTTTTAACCCACTCCCTGGATGCATTTTCCCTCCTGCATTTCCCTCTTTTTCCTGA 2024

Qy   1476 GTTCATACTAGAGAATCTGCACTATGTTTTTCCCCTTTTGTCTTGAGATGAAAGTTTAA 1535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2025 GTTCATACTAGAGAATCTGCACTATGTTTTTCCCCTTTTGTCTTGAGATGAAAGTTTAA 2084

Qy   1536 AATAATCCACTCTGTCATTTCCACTCTGAACATCCCAAGCTGTATCCCTGGCCTCTT 1595
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2085 AATAATCCACTCTGTCATTTCCACTCTGAACATCCCAAGCTGTATCCCTGGCCTCTT 2144

Qy   1596 TTCTCAGATCATGTTTCTTTTACTTGGGACCTAGAACTGGATTGGATTGCTCCTG 1655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2145 TTCTCAGATCATGTTTCTTTTACTTGGGACCTAGAACTGGATTGGATTGCTCCTG 2204

Qy   1656 ATCAGATGAGACCTTTGATTATTTGCCCCTTCCTTAGGACCTTACACTCCTGTCTTTCTT 1715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2205 ATCAGATGAGACCTTTGATTATTTGCCCCTTCCTTAGGACCTTACACTCCTGTCTTTCTT 2264

Qy   1716 TGACTTGCCTTTTGTTCTTTCCTTCATCCTTCAGTCCCCTTCCTCATGCAGTAGTCATTG 1775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2265 TGACTTGCCTTTTGTTCTTTCCTTCATCCTTCAGTCCCCTTCCTCATGCAGTAGTCATTG 2324
```

```
Qy   1776 CTAGGTAGAGAGGTATGTCCTTTTATGTAAATGCGACCACCGCATTTAGTATTACATAAAACTTTC 1835
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2325 CTAGGTAGAGAGGTATGTCCTTTTATGTAAATGCGACCACCGCATTTAGTATTACATAAAACTTTC 2384

Qy   1836 TTTTAACAATCGTGCATAGTACATGCTCGTCTGTTCCATTGAGATTTGACAGAGGTT 1895
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2385 TTTTAACAATCGTGCATAGTACATGCTCGTCTGTTCCATTTAGAGATTTGACAGAGGTT 2444

Qy   1896 TCAGTTTAGTATACTCAAATCTTATTTTAGTGCTTGGGAAATCAATTCAGAATATCACAT 1955
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2445 TCAGTTTAGTATACTCAAATCTTATTTTAGTGCTTGGGAAATCAATTCAGAATATCACAT 2504

Qy   1956 CCTTCCAATTCTCTCTACTCCAAATTGCTGGGAAACTCTCATGTTACTAACTTTGTTGTC 2015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2505 CCTTCCAATTCTCTCTACTCCAAATTGCTGGGAAACTCTCATGTTACTAACTTTGTTGTC 2564

Qy   2016 TCTAACTCTGCCATCTGGTTTCCCATCCTTCTTCCTCATGTACGTGTCCTA 2075
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2565 TCTAACTCTGCCATCTGGTTTCCCATCCTTCTTCCTCATGTACGTGTCCTA 2624

Qy   2076 ATATTAGCGTTGGTTGGAGATTTCAGTGCGTCCAATATTCCTTCCCTCGTTGGCCTTT 2135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2625 ATATTAGCGTTGGTTGGAGATTTCAGTGCGTCCAATATTCCTTCCCTCGTTGGCCTTT 2684

Qy   2136 CCTGAGATAATCCACTAAGAATATTTTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGAA 2195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2685 CCTGAGATAATCCACTAAGAATATTTTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGAA 2744

Qy   2196 ATTATCAACTGTGCACAAGGAAAAAATAGATATGTGAAAGTTCACGTAAATTCCTCA 2255
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2745 ATTATCAACTGTGCACAAGGAAAAAATAGATATGTGAAAGTTCACGTAAATTCCTCA 2804

Qy   2256 CATCACAGAAGATTAAAATTCAGAAAGGAGAAAAACAGACACAAAGAGAAGTATCTAAGA 2315
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2805 CATCACAGAAGATTAAAATTCAGAAAGGAGAAAAACAGACACAAAGAGAAGTATCTAAGA 2864

Qy   2316 CCAAAGGGATGTGTTTTATTAATGTCTAGAGTGAAGAAATGCATAGAACATTGTAGTACT 2375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2865 CCAAAGGGATGTGTTTTATTAATGTCTAGAGTGAAGAAATGCATAGAACATTGTAGTACT 2924

Qy   2376 TGTAAATAACTAGAAATAACATGATTTAGTCATAATTGTGAAAAATAATAATTTTTC 2435
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2925 TGTAAATAACTAGAAATAACATGATTTAGTCATAATTGTGAAAAATAATAATTTTTC 2984

Qy   2436 TTGGATTTATGTTCTGTATCTGTGAAAAAATAAATTCTTATAAAACTC 2484
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   2985 TTGGATTTATGTTCTGTATCTGTGAAAAAATAAATTCTTATAAAACTC 3033
```

```
RESULT 2
AC104837
LOCUS       AC104837     151265 bp    DNA     linear   PRI 19-JAN-2002
DEFINITION  Homo sapiens chromosome 1 clone RP4-641G12, complete sequence.
ACCESSION   AC104837 AL157947
VERSION     AC104837.2  GI:18249998
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151265)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151265)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 151265)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
```

TITLE Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.

JOURNAL Direct Submission
Submitted (19-JAN-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

COMMENT On Jan 19, 2002 this sequence version replaced gi:17976475.
-------- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchtgs@u.washington.edu
Drafting Center: SC
-------- Project Information
Center project name: chr-1
Center clone name: RP4-641G12 (sc0195)
-------- Summary Statistics
Sequencing vector: plasmid; L08752; 31% of reads
Sequencing vector: plasmid; 69% of reads
Chemistry: Dye-terminator ET; 75% of reads
Chemistry: Dye-terminator Big Dye; 25% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151256 bases at least Q40
Consensus quality: 151265 bases at least Q30
Consensus quality: 151265 bases at least Q20
Insert size: 151265; sum-of-contigs
Quality coverage: 13.7x in Q20 bases; sum-of-contigs

Overlapping Sequences:
5': RP11-369023 (UWGC:sc0385) AL354946
3': RP11-143H12 (UWGC:sc0641) AC099674

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

| EcoRI | | HindIII | | BglII | |
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
10449	10293	4702	4660	7487	7676
2184	2175	449	<800	5671	5775
8065	8212	512	<800	4181	4233
486	<800	2814	2820	7761	7676
106	<800	1247	1297	5168	5105
4682	4705	4695	4660	1735	1698
606	<800	3373	3381	529	<800
4198	4225	5592	5541	643	<800
3725	3694	2918	2948	1685	1698
5850	6047	4933	4843	981	921
418	<800	12634	12485	1768	1698
575	<800	1681	1612	3918	3844
337	<800	2417	2408	1954	1985
290	<800	765	781	144	<800
1493	1463	5897	5894	1072	1072
1375	1373	2196	2178	83	<800
777	786	1337	1297	2311	2150
2880	2953	671	<800	5786	5775
998	985	784	781	125	<800
119	<800	499	<800	5168	5105
1401	1373	6819	6736	2370	2369
305	<800	2344	2284	6218	6155
1953	1961	15	<800	1401	1387
8699	8762	73	<800	3103	3162
918	985	9	<800	1596	1534
6245	6047	15	<800	2183	2150
1288	1264	4408	4390	1008	1072
4012	3971	507	<800	2306	2150
2789	2842	6659	6736	6663	6646
109	<800	592	<800	2485	2511
791	786	4112	4049	1547	1534
1112	1093	2133	2178	3149	3162
1445	1373	4311	4257	4483	4486
3016	2953	1158	1195	448	<800
7827	7803	7399	7536	3382	3416
1713	1686	211	<800	1229	1237
881	890	3006	2948	6669	6646
5148	5107	6064	6115	6227	6155
2540	2582	1785	1863	2346	2369
407	<800	5272	5233	2369	2369
82	<800	6135	6115	3876	3844

```
Qy   1656  ATCAGATGAGACCTTGATTATTTGCCCTTCCTTAGGACCTTACACTCCTGTCTTTCTT  1715
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59575  ATCAGATGAGACCTTGATTATTTGCCCTTCCTTAGGACCTTACACTCCTGTCTTTCTT  59634

Qy   1716  TGACTTGCCTTTTGTTCTTCCTTCATCTTAGTCCTTCATGTCCAGTATGGTCATTG    1775
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59635  TGACTTGCCTTTTGTTCTTCCTTCATCTTAGTCCTTCATGTCCAGTATGGTCATTG    59694

Qy   1776  CTAGGTAGAGGTATGTCCTTTTATGTAATGGCCACCGCATTAGTATTACATAAACTTTC  1835
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59695  CTAGGTAGAGGTATGTCCTTTTATGTAATGGCCACCGCATTAGTATTACATAAACTTTC  59754

Qy   1836  TTTTAACAAATCTGTGCATAGTACATGCTGTCTGTTCCATTTAGAGATTTGACAGAGGTT 1895
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59755  TTTTAACAAATCTGTGCATAGTACATGCTGTCTGTTCCATTTAGAGATTTGACAGAGGTT 59814

Qy   1896  TCAGTTTAGTATACTCAAATCTTATTTAGTGCTTGGGAAATCAATTCAGAATATCACAT  1955
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59815  TCAGTTTAGTATACTCAAATCTTATTTAGTGCTTGGGAAATCAATTCAGAATATCACAT  59874

Qy   1956  CCTCTCCAATTCTCTCTCTACTCAAATTGCTGGGAAACTCTCATGTTACTAACTTTGTTGC 2015
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59875  CCTCTCCAATTCTCTCTCTACTCAAATTGCTGGGAAACTCTCATGTTACTAACTTTGTTGC 59934

Qy   2016  TCTAACTCTGCCATCTTGGTTTCCCCATCCCCTTCAGTGGTCCCAAATTCCTCTCC      2075
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59935  TCTAACTCTGCCATCTTGGTTTCCCATCCCCTTCAGTGGTCCCAAATTCCTCTCCTA      59994

Qy   2076  ATATTAGCGTTGGTTGGAGATTTCAGTGCGTCCAAATTCCTCTTCCTTCTTCAGGAATCCA  2135
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59995  ATATTAGCGTTGGTTGGAGATTTCAGTGCGTCCAAATTCCTCTTCCTTCTTCAGGAATCCA  60054

Qy   2136  CCTGAGATAATCCACTAAGAATATTTGTGTTCTTTCTTCAGGGAATCTAAGGAGGAA  2195
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  60055  CCTGAGATAATCCACTAAGAATATTTGTGTTCTTTCTTCAGGGAATCTAAGGAGGAA  60114

Qy   2196  ATTATCAACTGTGCACAAGGAGAAAAATAGATATTTCACGTAAATTCCTCA       2255
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db  60115  ATTATCAACTGTGCACAAGGAGAAAAATAGATATTTCACGTAAATTCCTCA       60174

Qy   2256  CATCACAGAAGATTAAATTCAGAAAGGAGAAAACACAGACCAAAGAGAAGTATCTAAGA  2315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  60175  CATCACAGAAGATTAAAATTCAGAAAGGAGAAAACACAGACCAAAGAGAAGTATCTAAGA  60234

Qy   2316  CCAAAGGGATGGTGTTTTATTAAATGTCTAGATGGAGAAATGCATAGAACATTGTAGTACT 2375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  60235  CCAAAGGGATGGTGTTTTATTAAATGTCTAGATGGAGAAATGCATAGAACATTGTAGTACT 60294

Qy   2376  TGTAAATAACTAGAAATAACATGATTTAGTCATAATTGTGAAAAATAATAATAATTTTTC  2435
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  60295  TGTAAATAACTAGAAATAACATGATTTAGTCATAATTGTGAAAAATAATAATAATTTTTC  60354

Qy   2436  TTGGATTTATGTTCTGTATCTGTGAAAAATAAATTTCTTATAAAACTC  2484
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db  60355  TTGGATTTATGTTCTGTATCTGTGAAAAATAAATTTCTTATAAAACTC  60403
```

RESULT 3
AX285019

```
LOCUS       AX285019                2527 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 824 from Patent WO0179556.
ACCESSION   AX285019
VERSION     AX285019.1  GI:17045707
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Lillie,J., Brown,J.L., Bolt,A. and van Huffel,C.
  TITLE     Novel genes, compositions and methods for the identification,
            assessment, prevention, and therapy of human cancers
```

```
Qy   6888                                                           1072
           ----------------------------------
Db   2901                                                           921

Qy   2027                                                           1985
Db   4831                                                           2750

Qy     78                                                           2750
Db   1185                                                           <800

Qy   3112                                                           2891
Db    144                                                           2750

Qy    402                                                           2891
Db    374                                                           921

Qy   2413                                                           8244
Db   2477                                                           1985

Qy   6449                                                           <800
Db   3465                                                           2369

Qy    801                                                           5105
Db    103                                                           763
```

Query Match   55.1%; Score 1369; DB 9; Length 151265;
Best Local Similarity 100.0%; Pred. No. 2.3e-260;
Matches 1369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1116  AGCTAGAGAGGAAGTCCAAAGAAAAACTTGGATTTGCTCTTTCTGACACATCCGGTTGGTTAGCA  1175
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59035  AGCTAGAGAGGAAGTCCAAAGAAAAACTTGGATTTGCTCTTTCTGACACATCCGGTTGGTTAGCA  59094

Qy   1176  ATTATTCCCTCGAGTGGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTTGCTCTGA  1235
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59095  ATTATTCCCTCGAGTGGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTTGCTCTGA  59154

Qy   1236  GACGAATGCTATGGGCTGCAGATGACTTCCTTAGAGGATTTGCCTTTGAGCAAATAGGTA  1295
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59155  GACGAATGCTATGGGCTGCAGATGACTTCCTTAGAGGATTTGCCTTTGAGCAAATAGGTA  59214

Qy   1296  GATGGTTTGGTGGTGTGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGGA  1355
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59215  GATGGTTTGGTGGTGTGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGGA  59274

Qy   1356  TCTATTAAATACCTGCAGCTCTCTGTCTTTTGTGGGTTGTTGCCCTGATTAGTTCT  1415
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59275  TCTATTAAATACCTGCAGCTCTCTGTCTTTTGTGGGTTGTTGCCCTGATTAGTTCT  59334

Qy   1416  GCTTTTTAACCACTCCCTGGATGCATTTTCCCCTCCGCATTTCCCTCTTTTCCTGGA  1475
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59335  GCTTTTTAACCACTCCCTGGATGCATTTTCCCCTCCGCATTTCCCTCTTTTCCTGGA  59394

Qy   1476  GTTCATACTAGAGAATCTGCACTACTATGTGTTTTGTCTTGAGATGAAAGTTTTAA  1535
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59395  GTTCATACTAGAGAATCTGCACTACTATGTGTTTTGTCTTGAGATGAAAGTTTTAA  59454

Qy   1536  AATAATCCACCTCGTCATTTCCACTCTCTGAACATCCAAGCTGTATCCCTGGCCTCTT  1595
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59455  AATAATCCACCTCGTCATTTCCACTCTCTGAACATCCAAGCTGTATCCCTGGCCTCTT  59514

Qy   1596  TTCTCAGACTATGTTTCTTTACTTGGGACCTAGAACTGGATTGGCATTGGCATTGCTCCTG  1655
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  59515  TTCTCAGACTATGTTTCTTTACTTGGGACCTAGAACTGGATTGGCATTGGCATTGCTCCTG  59574
```

```
JOURNAL   Patent: WO 0179555-A 824 25-OCT-2001;
          Millennium Predictive Medicine, Inc. (US)
FEATURES            Location/Qualifiers
     source         1..2527
                    /organism="Homo sapiens"
                    /mol_type="unassigned DNA"
                    /db_xref="taxon:9606"
ORIGIN

Query Match      49.7%;  Score 1235.2;  DB 6;  Length 2527;
Best Local Similarity   98.6%;  Pred. No. 1e-233;
Matches 1298;  Mismatches 14;  Conservative 0;  Indels 5;  Gaps 5;

Qy  1116  AGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTGACATCTCGGTGGTTAGCA  1175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1170  AGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTGACATCTCGGTGGTTAGCA  1229

Qy  1176  ATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTCTGCTCTGA  1235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1230  ATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTCTGCTCTGA  1289

Qy  1236  GACGAATGCTATGGGCTGCAGATGACTTCTTAGAGATTTGCCTTTGAGCAAATAGGTA  1295
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1290  GACGAATGCTATGGGCTGCAGATGACTTCTTAGAGATTTGCCTTTGAGCAAATAGGTA  1349

Qy  1296  GATGGTTGGTGTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGGA  1355
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1350  GATGGTTGGTGTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGGA  1409

Qy  1356  TCTATTAAATACCTGGCAGCTCTCTGTCTTTTTGTGGGTTGTTGCCCTGTGATTAGTTCT  1415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1410  TCTATTAAATACCTGGCAGCTCTCTGTCTTTTTGTGGGTTGTTGCCCTGTGATTAGTTCT  1469

Qy  1416  GCTTTTTAACCCACTCCCTGGATGCATTTTCCCTCCTTGCATTTCCCTCTTTTCCTGGA  1475
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1470  GCTTTTTAACCCACTCCCTGGATGCATTTTCCCTCCTTGCATTTCCCTCTTTTCCTGGA  1529

Qy  1476  GTTCATACTAGAGAATCTGCACTATGTTTTCCCTTTTGTCTTGAGATGAAAGTTTTAA  1535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1530  GTTCATACTAGAGAATCTGCACTATGTTTTCCCTTTTGTCTTGAGATGAAAGTTTTAA  1589

Qy  1536  AATAAATCCACCTCGTCATTCCACTCCTCGAACATCCCAAGCTGTATCCCTGGCCTCTT  1595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1590  AATAAATCCACCTCGTCATTCCACTCCTCGAACATCCCAAGCTGTATCCCTGGCCTCTT  1649

Qy  1596  TTCTCAGACTATGTTCTTTACTTGGGACCTAGAACTGGATTGGCATTGCTCCTG  1655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1650  TTCTCAGACTATGTTCTTTACTTGGGACCTAGAACTGGATTGGCATTGCTCCTG  1709

Qy  1656  ATCAGATAGACCTTTGATTATTTGCCCCTTCCTTAGGACCTTACACTCCTGTCTTTCTT  1715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1710  ATCAGATAGACCTTTGATTATTTGCCCCTTCCTTAGGACCTTACACTCCTGTCTTTCTT  1769

Qy  1716  TGACTTGCCTTTTGTTCTTTCCTTCATCTTAGTCCCTCTTCAGTATGGTCATTG  1775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1770  TGACTTGCCTTTTGTTCTTTCCTTCATCTTAGTCCCTCTTCAGTATGGTCATTG  1829

Qy  1776  CTAGGTAGAGGTATGTCCTTTATGTAATGGCACCGCATTTAGTAATTACATAAACTTTC  1835
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1830  CTAGGTAGAGGTATGTCCTTTATGTAATGGCACCGCATTTAGTAATTACATAAACTTTC  1889

Qy  1836  TTTTAACAATCTGTCATAGTACATGCTGCTCTGTTCCATTGAGATTGACAGAGGTT  1895
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1890  TTTTAACAATCTGTCATAGTACATGCTGCTCTGTTCCATTGAGATTGACAGAGGTT  1949

Qy  1896  TCAGTTTAGTATACTCAAATCTTATTTAGTGCTTGGGAAATCAATTCAGAATATCACAT  1955
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1950  TCAGTTTAGTATACTCAAATCTTATTTAGTGCTTGGGAAATCAATTCAGAATATCACAT  2009

Qy  1956  CCTCTCCAATTCTCTCTTACTCAAATTGCTGGGAAACTCTCATGTTACTAACTTTGTTGC  2015
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2010  CCTCTCCAATTCTCTCTTACTCAAATTGCTGGGAAACTCTCATGTTACTAACTTTGTTGC  2069
```

```
Qy  2016  TCTAACTCTGCCATCTTGGTTTCCCATCCTTCTTCTCTCATGGTAGGTGTGCTCCTA  2075
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2070  TCTAACTCTGCCATCTTGGTTTCCCATCCTTCTTCTCTCATGGTAGGTGTGCTCCTA  2129

Qy  2076  ATATTAGGCGTTGGTTGAGATTTCAGTGGTCCAATATTCCTCTTCCTCCGTGTTGCCTTT  2135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2130  ATATTAGGCGTTGGTTGAGATTTCAGTGGTCCAATATTCCTCTTCCTCCGTGTTGCCTTT  2189

Qy  2136  CCTGAGATAATCCACTAAGAAATATTTTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGAA  2195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2190  CCTGAGATAATCCACTAAGAAATATTTTGTGTTTCTTTTCTCAGGGAATCTAAGGGAGGAA  2249

Qy  2196  ATTATCAACTGTGCACAAGGAAAAAAATAGATGTGAAAGGTTCACGTAAATTCCTCA  2255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2250  ATTATCAACTGTGCACAAGGAAAAAAATAGATGTGAAAGGTTCACGTAAATTCCTCA  2309

Qy  2256  CATCACAGAGATTAAAATTCAGAAGAGAGAAAACACAGACCAA-AGAGAAGTATCTAAG  2314
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  2310  CATCACAGAGATTAAAATTCAGAAGAGAGAAAACACAGACCAAGAGAGAAGTATCTAAG  2369

Qy  2315  A-CCAAAGGGATGTGTTTATTAAT-GTCTAGGATGAAGAAATGCATA-GAACATTGTA-  2370
          | ||||||||||||||||||||||| |||||||||||||||||||||| |||||||||
Db  2370  ACCCAAAGGGATGTGTTTTATTAATGGTCTAGGATGGATGAAGAAATGCATAGGAACATTGTAG  2429

Qy  2371  GTACTTGTAAATAACTAGAAATAACATGATTTAGTCATATTGTGAAAATAATAAT  2427
          |||||||||||||||| |||||||||| |||||| ||  |||||||       |
Db  2430  GTACTTGTAAATAACCTAGAAATAACCTGGATTTGGGCCCATAATGTGAAAAAAAAANT  2486


RESULT 4
BC022870
LOCUS       BC022870                1714 bp    mRNA    linear   PRI 04-OCT-2003
DEFINITION  Homo sapiens interferon-induced protein 44, mRNA (cDNA clone
            MGC:24007 IMAGE:4071404), complete cds.
ACCESSION   BC022870
VERSION     BC022870.1  GI:18605553
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1714)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1714)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
```

COMMENT   Contact: MGC help desk.
                 Email: cgapbs-r@mail.nih.gov
                 Tissue Procurement: CLONTECH
                 cDNA Library Preparation: CLONTECH Laboratories, Inc.
                 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                 DNA Sequencing by: Institute for Systems Biology
                 http://www.systemsbiology.org
                 contact: amadan@systemsbiology.org
                 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
                 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

                 Clone distribution: MGC clone distribution information can be found
                 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
                 Series: IRAL Plate: 32 Row: n Column: 11
                 This clone was selected for full length sequencing because it
                 passed the following selection criteria: matched mRNA gi: 21361309.

FEATURES   Location/Qualifiers
     source   1..1714
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /clone="MGC:24007 IMAGE:4071404"
                 /tissue_type="Lung"
                 /clone_lib="NIH_MGC_77"
                 /lab_host="DH10B"
                 /note="Vector: pDNR-LIB"
     gene   1..1714
                 /gene="IFI44"
                 /note="synonyms: MTAP44, p44"
                 /db_xref="LocusID:10561"
     CDS   90..1424
                 /codon_start=1
                 /product="interferon-induced, hepatitis C-associated
                 microtubular aggregat"
                 /protein_id="AAH2870.1"
                 /db_xref="GI:18605554"
                 /db_xref="LocusID:10561"
                 /translation="MAVTTRLTRLHBKILQNHFGGKRLSLLYKGSVHGFRNGVLLDRC
                 CNQGPTLTVIYSEDHIGAYAEESYQEGKYASIILFALQDTKISEWKLGLCTPETLFC
                 CDVTKYNSPTNFQIDGRNRKVIMDLKTMENLGLAQNCTISIQDYEVFRCEDSLDERKI
                 KGVIELRKSLLSALRTYEPYGSLVQQIRILLGPIGAGKSSFFNSVRSVFQGHVTHQA
                 LVGTNTTGISEKYRTYSIRDGKDGKYLPFILCDSLGLSEKEGGLCRDDIFYILNGNIR
                 DRYQFNPMESIKLNHHDYIDSPSLKDRIHCVAFVFDASSIQYFSSQMIVKIKRIRREL
                 VNAGVVHVALLTHVDSMDLITKGDLIEIERCEPVRSKLEEVQRKLGFALSDISVVSNY
                 SSEWELDPVKDVLILSALRRMLWAADDFLEDLPFEQIGNLREEIINCAQGKK"
     misc_feature 141..302
                 /note="TLDc; Region: domain in TBC and LysM domain
                 containing proteins"
                 /db_xref="CDD:smart00584"

ORIGIN
Query Match   18.4%;   Score 456.6; DB 9; Length 1714;
Best Local Similarity   95.2%; Pred. No. 5.4e-80;
Matches   471;   Conservative   0;   Mismatches   24;   Indels   0;   Gaps   0;

Qy   801 AAATATTTTTGAATTAATAATCTGCTTATGTCACCTTACAGTTTAATCCCATGGAAT   860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   886 ACATATTCTATATCTTGAACGGTAACATTCGTGATAGATACCAGTTTAATCCCATGGAAT   945

Qy   861 CAATCAAATTAAATCATCATGACTACATTGATTCCCATCGCTGAAGGACAGAATTCATT   920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   946 CAATCAAATTAAATCATCATGACTACATTGATTCCCATCGCTGAAGGACAGAATTCATT   1005

Qy   921 GTGTGGCATTGTATTGATGCCAGCTCTATTCAATACTTCTCCTCAGATGATAGTAA   980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1006 GTGTGGCATTGTATTGATGCCAGCTCTATTCAATACTTCCCTCAGATGATAGTAA   1065

Qy   981 AGATCAAAAGAATTCGAAGGAGTTGGTAAACGCTGGTGGTGTGTACATGTGCTTTGCTA   1040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1066 AGATCAAAAGAATTCGAAGGAGTTGGTAAACGCTGGTGGTGTGTACATGTGCTTTGCTCA   1125

Qy   1041 CTCATGTGGATAGCCATGGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG   1100
          |||||||||||||||||||||||||||||||||

Db   1126 CTCATGTGGATAGCCATGGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG   1185

Qy   1101 AGCCTGTGAGGTCCAAGCTAGGAGGAGTCCAAAGAAAAACTTGGATTTGCTCTTTCTGACA   1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1186 AGCCTGTGAGGTCCAAGCTAGGAGAGTCCAAAGAAAAACTTGGATTTGCTCTTTCTGACA   1245

Qy   1161 TCTCGGTGGTTAGCAATTATCCTCTGAGTGGGAGCTGGACCCTGTAAAGGAATGTTCTAA   1220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1246 TCTCGGTGGTTAGCAATTATCCTCTGAGTGGGAGCTGGACCCTGTAAAGGAATGTTCTAA   1305

Qy   1221 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT   1280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1306 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT   1365

Qy   1281 TTGAGCAAATAGGTA   1295
          ||||||||||||||
Db   1366 TTGAGCAAATAGGGA   1380

RESULT 5
AX405991
LOCUS   AX405991   1679 bp   DNA   linear   PAT 14-JUN-2002
DEFINITION   Sequence 406 from Patent WO0222660.
ACCESSION   AX405991
VERSION   AX405991.1 GI:21439421
KEYWORDS   .
SOURCE   Homo sapiens (human)
   ORGANISM Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
   AUTHORS   Tang,Y.T., Liu,C., Zhou,P., Asundi,V., Zhang,J., Zhao,Q.A., Ren,F.,
             Xue,A.J., Yang,Y., Wehrman,T. and Drmanac,R.T.
   TITLE   Novel nucleic acids and polypeptides
   JOURNAL   Patent: WO 0222660-A 406 21-MAR-2002;
             HYSEQ, INC. (US)
FEATURES   Location/Qualifiers
     source   1..1679
                 /organism="Homo sapiens"
                 /mol_type="unassigned DNA"
                 /db_xref="taxon:9606"
     CDS   67..1401
                 /note="unnamed protein product"
                 /codon_start=1
                 /protein_id="CAD35098.1"
                 /db_xref="GI:21439422"
                 /db_xref="REMTREMBL:CAD35098"
                 /translation="MAVTTRLTTWLHEKKILQNHFGGKRLSLLYKGSVHGFRNGVLLDRC
                 CNQGPTLTVIYSEDHIIGAYABESYQEGKYASIILFALQDTKISEWKLGLCTPETLFC
                 CDVTKYNSPTNFQIDGRNRKVIMDLKTMENLGLAQNCTISIQDYEVFRCEDSLDERKI
                 KGVIELRKSLLSALRTYEPYGSLVQQIRILLGPIGAGKSSFFNSVRSVFQGHVTHQA
                 LVGTNTTGISEKYRTYSIRDGKDGKYLPFILCDSLGLSEKEGGLCRDDIFYILNGNIR
                 DRYQFNPMESIKLNHHDYIDSPSLKDRIHCVAFVFDASSIQYFSSQMIVKIKRIQREL
                 VNAGVVHVALLTHVDSMDLITKGDLIEIERCEPVRSKLEEVQRKLGFALSDISVVSNY
                 SSEWELDPVKDVLILSALRRMLWAADDFLEDLPFEQIGNLREEIINCAQGKK"
ORIGIN
Query Match   18.3%; Score 455; DB 6; Length 1679;
Best Local Similarity   94.9%; Pred. No. 1.1e-79;
Matches   470;   Conservative   0;   Mismatches   25;   Indels   0;   Gaps   0;

Qy   801 AAATATTTTTGAATTAATAATCTTGCTTTATGTCTACCTTACAGTTTAATCCCATGGAAT   860
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   863 ACATATTCTATATCTTGAACGGTAACATTCGTGATAGATACCAGTTTAATCCCATGGAAT   922

Qy   861 CAATCAAATTAAATCATCATGACTACATTGATTCCCATCGCTGAAGGACAGAATTCATT   920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   923 CAATCAAATTAAATCATCATGACTACATTGATTCCCATCGCTGAAGGACAGAATTCATT   982

Qy   921 GTGTGGCATTGTATTGATGATGCCAGCTCTATTCAATACTTCTCCTCTCAGATGATAGTAA   980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   983 GTGTGGCATTGTATTGATGATGCCAGCTCTATTCAATACTTCTCCTCTCCAGATGATAGTAA   1042

Qy 801 AAATATTTTTGAATAATATCTGCTTTATGTCTACCTTACAGTTTAATCCATGGAAT 860
        |||||||| || ||||||| || ||||||||||||||||||||||||||||
Db 524 ACATATCCTACATCTTGAACGGTAACATTCGTGATAGATACCAGTTTAATCCCATGGAAT 583

Qy 861 CAATCAAATTAAATCATCATGACTACATACATTGATTCCCATCGCTGAAGGACAGAATTCATT 920
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 584 CAATCAAATTAAATCATCATGACTACATACATTGATTCCCATCGCTGAAGGACAGAATTCATT 643

Qy 921 GTGTGCATTGTATTGATGCCAGCTCATTCATAATACTTCTCTCTCAGATGATAGTAA 980
        |||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db 644 GTGTGCATTGTATTGATGCCAGCTCATTCATAATACTTCTCCCTCTCAGATGATAGTAA 703

Qy 981 AGATCAAAAGAATTCGAAGGAGTTGGTAAACGCGGTGGTGTGGTACATGTGGCTTTGCTCA 1040
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 704 AGATCAAAAGAATTCGAAGGAGTTGGTAAACGCGGTGGTGTGGTACATGTGGCTTTGCTCA 763

Qy 1041 CTCATGTGGATAGCATGATTGATTACAAAAGTGACCTTATAGAAATAGAGAGATGTG 1100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 764 CTCATGTGGATAGCATGATTGATTACAAAAGGTGACCTTATACAAAAGGTGACCTTATAGAAATAGAGAGATGTG 823

Qy 1101 AGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACA 1160
        | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 824 TGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACA 883

Qy 1161 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA 1220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 884 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA 943

Qy 1221 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1280
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 944 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1003

Qy 1281 TTGAGCAAATAGGTA 1295
        |||||||||||| ||
Db 1004 TTGAGCAAATAGGGA 1018

RESULT 7
E01977
LOCUS   E01977   1062 bp   RNA   linear   PAT 29-SEP-1997
DEFINITION DNA encoding C-terminal peptide of type nonA nonB hepatitis
            specific antigen.
ACCESSION   E01977
VERSION   E01977.1 GI:2170225
KEYWORDS   JP 198912387-A/3.
SOURCE   Pan troglodytes (chimpanzee)
  ORGANISM Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 1062)
  AUTHORS   Shibui,T., Kurebayashi,R., Uesono,M., Teranishi,Y., Takahashi,K.,
            Nakanishi,S. and Kitamura,N.
  TITLE   MANIFESTATION VECTOR HAVING DNA CODING NON-A NON-B HEPATITIS
            SPECIFIC ANTIGEN, TRANSFORMANT AND PRODUCTION OF SAID ANTIGEN
  JOURNAL   Patent: JP 198912387-A 3 17-MAY-1989;
            MITSUBISHI KASEI CORP
COMMENT   OS   Pan troglodytes
            PN   JP 198912387-A/3
            PD   17-MAY-1989
            PF   10-NOV-1987 JP 1987283990
            PI   SHIBUI TATSURO, KUREBAYASHI RIE, UESONO MICHIRU, PI
            TERANISHI YUTAKA,
            PI   TAKAHASHI KAZUNOBU, NAKANISHI SHIGETADA, KITAMURA NAOMI PC
            C12N15/00,A61K39/29,C12N1/20,(C12P21/02,(C12P21/02,C12R1:19); CC
            strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: tissue_type=Liver;
            CC   *source: clone=PCDVCL-1;
            CC   Feature is identified by experimental;
            FH   Key   Location/Qualifiers
            FH

Qy 981 AGATCAAAAGAATTCGAAGGAGTTGGTAAACGCTGGTGTGGTGGTACATGTGGCTTTGCTCA 1040
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1043 AGATCAAAAGAATTCGAAGGAGTTGGTAAACGCTGGTGTGGTGGTACATGTGGCTTTGCTCA 1102

Qy 1041 CTCATGTGGATAGCATGATTGATTACAAAAGTGACCTTATAGAAATAGAGAGATGTG 1100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1103 CTCATGTGGATAGCATGATTGATTACAAAAGTGACCTTATAGAAATAGAGAGATGTG 1162

Qy 1101 AGCCTGTGAGGTCCAAGCTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGAAACTTGGATTTGCTCTTTCTGACA 1160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1163 AGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTCTGACA 1222

Qy 1161 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA 1220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1223 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA 1282

Qy 1221 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1280
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1283 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1342

Qy 1281 TTGAGCAAATAGGTA 1295
        |||||||||||| ||
Db 1343 TTGAGCAAATAGGGA 1357

RESULT 6
E01792
LOCUS   E01792   1062 bp   RNA   linear   PAT 29-SEP-1997
DEFINITION DNA sequence coding for non-A non-B hepatitis specific antigen
            protein.
ACCESSION   E01792
VERSION   E01792.1 GI:2170045
KEYWORDS   JP 1989002576-A/2.
SOURCE   Pan paniscus (pygmy chimpanzee)
  ORGANISM Pan paniscus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 1062)
  AUTHORS   Takahashi,K., Shibui,T., Uchida,M., Kurebayashi,R., Teranishi,Y.,
            Nakanishi,S. and Kitamura,N.
  TITLE   DNA FRAGMENT
  JOURNAL   Patent: JP 1989002576-A 2 06-JAN-1989;
            MITSUBISHI KASEI CORP
COMMENT   OS   Pan paniscus (Chimpanzee)
            PN   JP 1989002576-A/2
            PD   06-JAN-1989
            PF   04-JUN-1987 JP 1987140586
            PR   31-MAR-1987 JP 87P   78813
            PI   TAKAHASHI KAZUNOBU, SHIBUI TATSURO, UCHIDA MICHIRU, PI
            KUREBAYASHI RIE,
            PI   TERANISHI YUTAKA, NAKANISHI SHIGETADA, KITAMURA NAOMI PC
            C12N15/00,(C12N15/00,C12R1:91);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: tissue_type=Liver;
            FH   Key   Location/Qualifiers
            FH
            FT   CDS   1..1062
            FT   /product='Non-A non-B hepatitis specific FT
                                 antigen protein'.
FEATURES   Location/Qualifiers
     source   1..1062
                     /organism="Pan paniscus"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:9597"
ORIGIN

Query Match   18.1%; Score 448.6; DB 6; Length 1062;
Best Local Similarity   94.1%; Pred. No. 2.2e-78;
Matches 466; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
FT   misc_feature   1..1049
FT                  /note='DNA encoding C-terminal peptide of type
FT                  nonA nonB
FT                  hepatitis specific antigen'.
FEATURES     source    Location/Qualifiers
             1..1062
                  /organism="Pan troglodytes"
                  /mol_type="genomic RNA"
                  /db_xref="taxon:9598"
ORIGIN

Query Match   18.1%;   Score 448.6; DB 6; Length 1062;
Best Local Similarity  94.1%;  Pred. No. 2.2e-78;
Matches 466;  Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy  801 AAATATTTTTGAATTAATATCTTGCTTATGTCTACCTTACAGTTAATCCATGGAAT 860
        ||||||| |||||||||||    |   |||| ||||||||||||||||||
Db  524 ACATATCCTACATCTTGAACGGTAACATTCGTGATAGATACCAGTTAATCCATGGAAT 583

Qy  861 CAATCAAATTAAATCATCATGACTACATCATTGATTCCCATCGCTGAAGGACAGAATTCATT 920
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  584 CAATCAAATTAAATCATCATGACTACATCATTGATTCCCATCGCTGAAGGACAGAATTCATT 643

Qy  921 GTGTGGCATTTGTATTTGATGCCAGCTCATTCAATAACTTCTCCTCTCAGATGATAGTAA 980
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  644 GTGTGGCATTTGTATTTGATGCCAGCTCATTCAATAACTTCTCCTCTCAGATGATAGTAA 703

Qy  981 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGGCTTGCTCA 1040
        |||||||||||||||||||||||||||||||||||||||||||||
Db  704 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGGCTTGCTCA 763

Qy 1041 CTCATGTGGATAGCATGGATTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1100
        |||||||||||||||||||||||||||||||||||||||||||||
Db  764 CTCATGTGGATAGCATGGATTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG 823

Qy 1101 AGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTTCTGACA 1160
        ||||||||||||||||||||||||||||||||||||||||||||
Db  824 TGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTTCTGACA 883

Qy 1161 TCTCGGTGGTTAGCAATTATTCCTCGAGTGGGAGGTGGGACCCTGTAAAGGATGTTCTAA 1220
        |||||||||||||||||||||||||||||||||||||||||||||
Db  884 TCTCGGTGGTTAGCAATTATTCCTCGAGTGGGAGCTGGGACCCTGTAAAGGATGTTCTAA 943

Qy 1221 TTCTTTCGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1280
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  944 TTCTTTCGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1003

Qy 1281 TTGAGCAAATAGGTA 1295
        ||||||||||||||
Db 1004 TTGAGCAAATAGGGA 1018
```

RESULT 8
E01791
```
LOCUS       E01791    1332 bp    DNA    linear   PAT 29-SEP-1997
DEFINITION  DNA sequence coding for non-A non-B hepatitis specific antigen
            protein.
ACCESSION   E01791
VERSION     E01791.1  GI:2170044
KEYWORDS    JP 1989002576-A/1.
SOURCE      Pan paniscus (pygmy chimpanzee)
  ORGANISM  Pan paniscus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 1332)
  AUTHORS   Takahashi,K., Shibui,T., Uchida,M., Kurebayashi,R., Teranishi,Y.,
            Nakanishi,S. and Kitamura,N.
  TITLE     DNA FRAGMENT
  JOURNAL   Patent: JP 1989002576-A 1 06-JAN-1989;
            MITSUBISHI KASEI CORP
COMMENT     OS  Pan paniscus (Chimpanzee)
            PN  JP 1989002576-A/1
            PD  06-JAN-1989
```

```
PF  04-JUN-1987 JP 1987140586
PR  31-MAR-1987 JP 87P JP 87P    78313
PI  TAKAHASHI KAZUNOBU, SHIBUI TATSURO, UCHIDA MICHIRU, PI
    KUREBAYASHI RIE,
PI  TERANISHI YUTAKA, NAKANISHI SHIGETADA, KITAMURA NAOMI PC
    C12N15/00,(C12N15/00,C12R1:9);
CC  strandedness: Double;
CC  topology: Linear;
CC  hypothetical: No;
CC  anti-sense: No;
CC  *source: tissue_type=Liver;
FH  Key          Location/Qualifiers
FT  mat_peptide  1..1332
FT               /product='Non-A non-B hepatitis specific
                 antigen protein'.
FEATURES     source    Location/Qualifiers
             1..1332
                  /organism="Pan paniscus"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:9597"
ORIGIN

Query Match   18.1%;   Score 448.6; DB 6; Length 1332;
Best Local Similarity  94.1%;  Pred. No. 2.1e-78;
Matches 466;  Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy  801 AAATATTTTTGAATTAATATCTTGCTTATGTCTACCTTACAGTTAATCCATGGAAT 860
        ||||||| |||||||||||    |   |||| |||||||||||||||||||
Db  797 ACATATCCTACATCTTGAACGGTAACATTCGTGATAGATACCAGTTAATCCATGGAAT 856

Qy  861 CAATCAAATTAAATCATCATGACTACATCATTGATTCCCATCGCTGAAGGACAGAATTCATT 920
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  857 CAATCAAATTAAATCATCATGACTACATCATTGATTCCCATCGCTGAAGGACAGAATTCATT 916

Qy  921 GTGTGGCATTTGTATTTGATGCCAGCTCATTCAATAACTTCTCCTCTCAGATGATAGTAA 980
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  917 GTGTGGCATTTGTATTTGATGCCAGCTCATTCAATAACTTCTCCTCTCAGATGATAGTAA 976

Qy  981 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGGCTTGCTCA 1040
        |||||||||||||||||||||||||||||||||||||||||||||
Db  977 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGGCTTGCTCA 1036

Qy 1041 CTCATGTGGATAGCATGGATTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1100
        |||||||||||||||||||||||||||||||||||||||||||||
Db 1037 CTCATGTGGATAGCATGGATTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1096

Qy 1101 AGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTTCTGACA 1160
        ||||||||||||||||||||||||||||||||||||||||||
Db 1097 TGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTTTCTGACA 1156

Qy 1161 TCTCGGTGGTTAGCAATTATTCCTCGAGTGGGAGGTGGGACCCTGTAAAGGATGTTCTAA 1220
        |||||||||||||||||||||||||||||||||||||||||||||
Db 1157 TCTCGGTGGTTAGCAATTATTCCTCGAGTGGGAGCTGGGACCCTGTAAAGGATGTTCTAA 1216

Qy 1221 TTCTTTCGCTCTGAGACGAATGGCTGCAGATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1280
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 1217 TTCTTTCGCTCTGAGACGAATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1276

Qy 1281 TTGAGCAAATAGGTA 1295
        ||||||||||||||
Db 1277 TTGAGCAAATAGGGA 1291
```

RESULT 9
E01975
```
LOCUS       E01975    1332 bp    RNA    linear   PAT 29-SEP-1997
DEFINITION  DNA encoding type non-A non-B hepatitis specific antigen.
ACCESSION   E01975
VERSION     E01975.1  GI:2170223
KEYWORDS    JP 198912437-A/1.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
```

```
Db   1277 TTGAGCAAATAGGGA 1291
          |||||||||||||||
```

RESULT 10
CHPP44

```
LOCUS       CHPP44       1665 bp    mRNA    linear   PRI 21-DEC-2002
DEFINITION  Pan troglodytes gene for non-A non-B hepatitis-associated
            microtubular aggregates protein p44, complete cds.
ACCESSION   D90034
VERSION     D90034.1  GI:218575
KEYWORDS    .
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 1665)
  AUTHORS   Takahashi,K., Kitamura,N., Shibui,T., Kamizono,M., Matsui,R.,
            Yoshiyama,Y., Maeda,T., Kondo,J., Honda,J., Yamada,E.,
            Shimizu,Y.K., Teranishi,Y. and Nakanishi,S.
  TITLE     Cloning, sequencing and expression in Escherichia coli of cDNA for
            a non-A, non-B hepatitis-associated microtubular aggregates protein
  JOURNAL   J. Gen. Virol. 71 (Pt 9), 2005-2011 (1990)
   MEDLINE  91011346
   PUBMED   2170570
COMMENT     These data kindly submitted in computer readable form by:
            Yoshikazu Honda
            Mitsubishi Kasei Co.
            1000, Kamoshida-cho, Midori-ku
            Yokohama 227
            Japan
            Phone:  81-45-963-3454
            Fax:    81-45-963-3992.
FEATURES             Location/Qualifiers
     source          1..1665
                     /organism="Pan troglodytes"
                     /mol_type="mRNA"
                     /db_xref="taxon:9598"
                     /clone="pCDVCL-1"
                     /cell_type="hepatocyte"
                     /note="CH41"
     CDS             55..1389
                     /codon_start=1
                     /product="microtubular aggregates protein p44"
                     /protein_id="BAA14082.1"
                     /db_xref="GI:218576"
                     /translation="MAVTTRLTTWLHEKILQNHFGGKRLSLLYKGSVHGFHNGVLLDRC
                     CNQGPTLTVIYSEDHIIGAYAEBGYQERKYASIILFALQETKISEWKLGLYTPETLFC
                     CDVAKYNSPTNFQIDGRNRKVIMDLKTMENLGLAQNCTISIQDYEVFRCEDSLDERKI
                     KGVIELRKSLLSALRTYEPYGSLVQQTRILLLGPIGACKSSFFNSVRSVFQGHVTHQA
                     LVGTNTTGISEKXRTYSIRDGKDGKYLPFILCDSLGLSEKEGGLCMDDISYILNGNIR
                     DRYQFNPMESIKLNHHDYIDSPSLKDRIHCVAFVFDASSIEYFSSQMIVKIKIRREL
                     VNAGVHVALLTHVDSMDLITKGDLIEIERCVPVRSKLEEVQRKLGFALSDISVVSNY
                     SSEWELDVKDVLJLSALRRWLWAADDFLEDLPFEQIGNLREEIINCAQGKK"
     misc_feature    466..474
                     /note="put. N-glycosylation site"
     misc_feature    721..729
                     /note="put. N-glycosylation site"
     misc_feature    1225..1233
                     /note="put. N-glycosylation site"
     polyA_signal    1628..1633
ORIGIN
Query Match    18.1%;    Score 448.6;  DB 9;  Length 1665;
Best Local Similarity  94.1%;   Pred. No. 2.1e-78;
Matches 466;   Conservative 0;  Mismatches 29;   Indels 0;  Gaps 0;

Qy   801 AAATATTTTTGAATTAATATCTTGCTTATGTCCATTACCAGTTAATCCCATGGAAT 860
         |||      ||||    ||    |||| ||||||||||
Db   851 ACATATCCTACATCTTGAACGGTAACATTCCTGATAGATACCAGTTAATCCCATGGAAT 910
Qy   861 CAATCAAATTAAATCATCATGAGCACACATTGATTCCCCATCGCTGAAGGACAGAATTCATT 920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       CAATCAAATTAAATCATCATGAGCACACATTGATTCCCCATCGCTGAAGGACAGAATTCATT 920
```

REFERENCE
AUTHORS

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
            1  (bases 1 to 1332)
            Shibui,T., Kurebayashi,R., Uesono,M., Teranishi,Y., Takahashi,K.,
            Nakanishi,S. and Kitamura,N.
TITLE       MANIFESTATION VECTOR HAVING DNA CODING NON-A NON-B HEPATITIS
            SPECIFIC ANTIGEN, TRANSFORMANT AND PRODUCTION OF SAID ANTIGEN
JOURNAL     Patent: JP 1989124387-A 1 17-MAY-1989;
            MITSUBISHI KASEI CORP
COMMENT     OS   Pan troglodytes
            PN   JP 1989124387-A/1
            PD   17-MAY-1989
            PF   10-NOV-1987 JP 1987283990
            PI   SHIBUI TATSURO, KUREBAYASHI RIE, UESONO MICHIRU, PI
            TERANISHI YUTAKA,
            PI   TAKAHASHI KAZUNOBU, NAKANISHI SHIGETADA, KITAMURA NAOMI PC
            C12N15/00,A61K39/29,C12N1/20,C12P21/02,(C12P21/02,C12R1:19); CC
            strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: tissue type=Liver;
            CC   *source: clone=pCDVCL-I;
            CC   Feature is identified by experimental;
FH   Key             Location/Qualifiers
FH
FT   mat_peptide     1..1332
FT                   /product='type non-A non-B hepatitis specific
FT                   antigen'.
FEATURES             Location/Qualifiers
     source          1..1332
                     /organism="Pan troglodytes"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:9598"
ORIGIN
Query Match    18.1%;    Score 448.6;  DB 6;  Length 1332;
Best Local Similarity  94.1%;   Pred. No. 2.1e-78;
Matches 466;   Conservative 0;  Mismatches 29;   Indels 0;  Gaps 0;

Qy   801 AAATATTTTTGAATTAATATCTTGCTTATGTCCATTACCAGTTAATCCCATGGAAT 860
         |||      ||||    ||    |||||||||||||
Db   797 ACATATCCTACATCTTGAACGGTAACATTCGTGATAGATACCAGTTAATCCCATGGAAT 856
Qy   861 CAATCAAATTAAATCATCATGAGCACACATTGATTCCCCATCGCTGAAGGACAGAATTCATT 920
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   857 CAATCAAATTAAATCATCATGAGCACACATTGATTCCCCATCGCTGAAGGACAGAATTCATT 916
Qy   921 GTGTGGCATTGTATTGATGCCAGCTCTATTCAATACTTCTCCTCTCAGATGATAGTAA 980
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   917 GTGTGGCATTGTATTGATGCCAGCTCTATTCAATACTTCTCCTCTCAGATGATAGTAA 976
Qy   981 AGATCAAAAGAATTCGAAGGAGGAGTCCAAGCTGGTGTGGTACATGTGGCTTTGCTCA 1040
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   977 AGATCAAAAGAATTCGAAGGAGGAGTCCAAGCTGGTGTGGTACATGTGGCTTTGCTCA 1036
Qy  1041 CTCATGTGGATAGCATGGATTGATTACAAAAGGTGACCTTATAGAAATAGAGATGTG 1100
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1037 CTCATGTGGATAGCATGGATCGATTATTCCCTGAGTGGAGCTCGACCCTGTAAAGGATGTTCTAA 1096
Qy  1101 AGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTGACA 1160
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  1097 TGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTGACA 1156
Qy  1161 TCTCGGTGGTTAGCAATTATTCCCTCGAGTGGAGCTCGACCCTGTAAAGGATGTTCTAA 1220
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1157 TCTCGGTGGTTAGCAATTATTCCCTGAGTGGAGCTCGACCCTGTAAAGGATGTTCTAA 1216
Qy  1221 TTCTTTCTGCTCTGAGACGAATGCTAGGGCTGCAGATGACTTCTTAGAGATTTGCCTT 1280
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  1217 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT 1276
Qy  1281 TTGAGCAAATAGGTA 1295
```

```
Db   911  CAATCAAATTAAATCATCATGACTACATTGATTCCCCATGCTGAAGGACAGAATTCATT  970

Qy   921  GTGTGGCATTTGTATTTGATGCCAGCTCTATTCAATACTTCTCCTCCTCAGATGATAGTAA  980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   971  GTGTGGCATTTGTATTTGATGCCAGCTCTATTGAATACTTCTCCTCCTCAGATGATAGTAA  1030

Qy   981  AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGGTACATGTGGCTTTGCTCA  1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1031  AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGGTACATGTGGCTTTGCTCA  1090

Qy  1041  CTCATGTGGATAGCATGGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG  1100
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  1091  CTCATGTGGATAGCATGGATCTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG  1150

Qy  1101  AGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAAACTTGGATTTGCTCTTTCTGACA  1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1151  TGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAAACTTGGATTTGCTCTTTCTGACA  1210

Qy  1161  TCTCGGTGGTTAGCAATTATTCCTCGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA  1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1211  TCTCGGTGGTTAGCAATTATTCCTCGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA  1270

Qy  1221  TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT  1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1271  TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT  1330

Qy  1281  TTGAGCAAATAGGTA  1295
          |||||||||||||||
Db  1331  TTGAGCAAATAGGGA  1345
```

RESULT 11

LOCUS   E01793   1667 bp   RNA   linear   PAT 29-SEP-1997
DEFINITION DNA sequence coding for non-A non-B hepatitis specific antigen
            protein.
ACCESSION   E01793
VERSION   E01793.1 GI:2170046
KEYWORDS   JP 1989002576-A/3.
SOURCE   Pan paniscus (Chimpanzee)
  ORGANISM Pan paniscus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 1667)
  AUTHORS   Takahashi,K., Shibui,T., Uchida,M., Kurebayashi,R., Teranishi,Y.,
            Nakanishi,S. and Kitamura,N.
  TITLE   DNA FRAGMENT
  JOURNAL   Patent: JP 1989002576-A 3 06-JAN-1989;
            MITSUBISHI KASEI CORP
COMMENT   OS   Pan paniscus (Chimpanzee)
            PN   JP 1989002576-A/3.
            PD   06-JAN-1989
            PF   04-JUN-1987 JP 1987140586
            PR   31-MAR-1987 JP 87P   78313
            PI   TAKAHASHI KAZUNOBU, SHIBUI TATSURO, UCHIDA MICHIRU, PI
            KUREBAYASHI RIE,
            PI   TERANISHI YUTAKA, NAKANISHI SHIGETADA, KITAMURA NAOMI PC
            C12N15/00, (C12N15/00,C12R1:91);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   * source: tissue_type=Liver;
            FH   Key   Location/Qualifiers
            FH
            FT   5'UTR   1..56
            FT   CDS   57..1391
            FT   /product='Non-A non-B hepatitis specific   FT
            antigen protein'
            FT   3'UTR   1392..1667.
FEATURES   Location/Qualifiers
     source   1..1667
                     /organism="Pan paniscus"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:9597"

ORIGIN

Query Match   18.1%; Score 448.6; DB 6; Length 1667;
Best Local Similarity   94.1%; Pred. No. 2.1e-78;
Matches 466; Conservative   0; Mismatches 29; Indels   0; Gaps   0;

```
Qy   801  AAATATTTTTGAATTAATATCTTGCTTTATGTCTACCTTACAGTTTAATCCCATGGAAT   860
          |||||||||||||||||||| ||||| |||| ||||||  |||||||||||||||||||
Db   853  ACATATCCTACATCTTGAACGTAACATTCGTGATAGATACCAGTTTAATCCCATGGAAT   912

Qy   861  CAATCAAATTAAATCATCATGACTACATTGATTCCCATGCTGAAGGACAGAATTCATT   920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   913  CAATCAAATTAAATCATCATGACTACATTGATTCCCATGCTGAAGGACAGAATTCATT   972

Qy   921  GTGTGGCATTTGTATTTGATGCCAGCTCTATTCAATACTTCTCCTCCTCAGATGATAGTAA   980
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   973  GTGTGGCATTTGTATTTGATGCCAGCTCTATTGAATACTTCTCCTCCTCAGATGATAGTAA  1032

Qy   981  AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGGTACATGTGGCTTTGCTCA  1040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1033  AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGGTACATGTGGCTTTGCTCA  1092

Qy  1041  CTCATGTGGATAGCATGGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG  1100
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  1093  CTCATGTGGATAGCATGGATCTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG  1152

Qy  1101  AGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAAACTTGGATTTGCTCTTTCTGACA  1160
          | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1153  TGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAAACTTGGATTTGCTCTTTCTGACA  1212

Qy  1161  TCTCGGTGGTTAGCAATTATTCCTCGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA  1220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1213  TCTCGGTGGTTAGCAATTATTCCTCGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAA  1272

Qy  1221  TTCTTTCTGCTCTGAGACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT  1280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1273  TTCTTTCTGCTCTGAGACCAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTT  1332

Qy  1281  TTGAGCAAATAGGTA  1295
          ||||||||||| |||
Db  1333  TTGAGCAAATAGGGA  1347
```

RESULT 12

LOCUS   E01978   1667 bp   RNA   linear   PAT 29-SEP-1997
DEFINITION DNA encoding type nonA nonB hepatitis specific antigen.
ACCESSION   E01978
VERSION   E01978.1 GI:2170226
KEYWORDS   JP 1989124387-A/4.
SOURCE   Pan troglodytes (chimpanzee)
  ORGANISM Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 1667)
  AUTHORS   Shibui,T., Kurebayashi,R., Uesono,M., Teranishi,Y., Takahashi,K.,
            Nakanishi,S. and Kitamura,N.
  TITLE   MANIFESTATION VECTOR HAVING DNA CODING NON-A NON-B HEPATITIS
            SPECIFIC ANTIGEN, TRANSFORMANT AND PRODUCTION OF SAID ANTIGEN
  JOURNAL   Patent: JP 1989124387-A 4 17-MAY-1989;
            MITSUBISHI KASEI CORP
COMMENT   OS   Pan troglodytes
            PN   JP 1989124387-A/4
            PD   17-MAY-1989
            PF   10-NOV-1987 JP 1987283990
            PP   SHIBUI TATSURO, KUREBAYASHI RIE, UESONO MICHIRU, PI
            TERANISHI YUTAKA,
            PI   TAKAHASHI KAZUNOBU, NAKANISHI SHIGETADA, KITAMURA NAOMI PC
            C12N15/00,A61K39/29,C12N1/20,C12P21/02,(C12P21/02,C12R1:19); CC
            strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;

```
CC      *source: tissue_type=Liver;
CC      *source: clone=pCDVCL-I;
CC      Feature is identified by experimental;
FH      Key             Location/Qualifiers
FH
FT      misc_feature    1..1667
FT                      /note='DNA containing complete sequence of
FT                      type norA nonB
                        hepatitis specific antigen'.
FEATURES              Location/Qualifiers
     source           1..1667
                      /organism="Pan troglodytes"
                      /mol_type="genomic RNA"
                      /db_xref="taxon:9598"
ORIGIN

Query Match     18.1%;  Score 448.6; DB 6; Length 1667;
Best Local Similarity   94.1%;  Pred. No. 2.1e-78;
Matches  466;  Conservative  0;  Mismatches  29;  Indels  0;  Gaps  0;

Qy    801 AAATATTTTTGAATTAATATCTTGCTTATGTCTACCTTACAGTTAATCCCATGGAAT 860
          |||| || ||||||||||||||||||| || ||  ||  |||||||||||||||||
Db    853 ACATATCCTACATCTTGAACGGTAACATTCGTGATAGTACCAGTTTAATCCCATGGAAT 912

Qy    861 CAATCAAATTAAATCATCATGACTACATTGATTCCCATCGCTGAAGGACAGAATTCATT 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    913 CAATCAAATTAAATCATCATGACTACATTGATTCCCATCGCTGAAGGACAGAATTCATT 972

Qy    921 GTGTGGCATTTGTATTTGATGCCAGCTCTATTCAATACTTCTCCTCTCAGATGATAGTAA 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    973 GTGTGGCATTTGTATTTGATGCCAGCTCTATTCAATACTTCTCCTCTCAGATGATAGTAA 1032

Qy    981 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGGTACATGTGGCTTTGCTCA 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1033 AGATCAAAAGAATTCGAAGGGAGTTGGTAAACGCTGGTGTGGTACATGTGGCTTTGCTCA 1092

Qy   1041 CTCATGTGGATAGCAATGGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1093 CTCATGTGGATAGCAATGGATCTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1152

Qy   1101 AGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTTCTGACA 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1153 TGCCTGTGAGGTCCAAGCTAGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTTCTGACA 1212

Qy   1161 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGAGCTGGACCCTGTAAAGATGTTCTAA 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1213 TCTCGGTGGTTAGCAATTATTCCTCTGAGTGGAGCTGGACCCTGTAAAGATGTTCTAA 1272

Qy   1221 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTCTCGAGATGACTTCTTAGAGGATTGCCTT 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1273 TTCTTTCTGCTCTGAGACGAATGCTATGGGCTCTCGAGATGACTTCTTAGAGGATTGCCTT 1332

Qy   1281 TTGAGCAAATAGGTA 1295
          ||||||||||||||
Db   1333 TTGAGCAAATAGGGA 1347
```

RESULT 13
```
LOCUS       AX284603                 567 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 408 from Patent WO0179556.
ACCESSION   AX284603
VERSION     AX284603.1  GI:17045291
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Lillie,J., Brown,J.L., Bolt,A. and van Huffel,C.
  TITLE     Novel genes, compositions and methods for the identification,
            assessment, prevention, and therapy of human cancers
  JOURNAL   Patent: WO 0179556-A 408 25-OCT-2001;
```

Millennium Predictive Medicine, Inc. (US)
```
FEATURES             Location/Qualifiers
     source          1..567
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match     16.7%;  Score 415.4; DB 6; Length 567;
Best Local Similarity   98.4%;  Pred. No. 8.6e-72;
Matches  419;  Conservative  0;  Mismatches  7;  Indels  0;  Gaps  0;

Qy   1116 AGCTAGGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTTCTGACAATCTCGGTGGTTAGCA 1175
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142 AGCTAGGAGGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTTCTGACAATCTCGGTGGTTAGCA 201

Qy   1176 ATTATTCCTCTGAGTGGAGCTGGACCCTGTAAAGATGTTCTAATTCTTTCTGCTCTGA 1235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    202 ATTATTCCTCTGAGTGGAGCTGGACCCTGTAAAGATGTTCTAATTCTTTCTGCTCTGA 261

Qy   1236 GACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTGCCTTTTGAGCAAATAGGTA 1295
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262 GACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTGCCTTTTGAGCAAATAGGTA 321

Qy   1296 GATGGTTTGGTGGTGGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGGA 1355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322 GATGGTTTGGTGGTGGGAAGCTTGGAAGCGGTCAGGTAGTTGGCTACTTTCTGCTTGGA 381

Qy   1356 TCTATTAAATACCTGGCAGCTCTCGTCTTTGTGGGTTGTGCCCTGTGATTAGTTCT 1415
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    382 TCTATTAAATACCTGGCAGCTCTCGTCTTTNTGGGTTGTGCCCTGTGATTAGTTCT 441

Qy   1416 GCTTTTTAACCCACTCCCTGGATGCATTTTCCCTCTTCTTGCATTTTCCCTCTTCTTTCCTGGA 1475
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    442 GCTTTTTAACCCACTCCCTGGATGCATTTTCCCTCTTCTTGCATTTTCCCTCTTCTTTCCTGGA 501

Qy   1476 GTTCATACTAGAGAATCTGCACTACATGTTTTCCTTTTGTCTTGAGATGAAAGTTTTAA 1535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    502 AGTCATACTAGAGAATCTGCACTACATGTTTTCCTTTTGTCTTGAGATGAAAGTTTTAA 561

Qy   1536 AATAAT 1541
          ||||||
Db    562 AATAAT 567
```

RESULT 14
```
LOCUS       D2890898                 483 bp    DNA     linear   PRI 14-SEP-2002
DEFINITION  Homo sapiens gene for hepatitis C-associated microtubular aggregate
            protein p44, exon 9 and complete cds.
ACCESSION   D28915
VERSION     D28915.1  GI:598954
KEYWORDS    .
SEGMENT     8 of 8
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 49 to 483)
  AUTHORS   Kitamura,A., Takahashi,K., Okajima,A. and Kitamura,N.
  TITLE     Induction of the human gene for p44, a hepatitis-C-associated
            microtubular aggregate protein, by interferon-alpha/beta
  JOURNAL   Eur. J. Biochem. 224 (3), 877-883 (1994)
   MEDLINE   95010078
   PUBMED    7925411
REFERENCE   2  (bases 1 to 483)
  AUTHORS   Kitamura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-1994) Naomi Kitamura, Institute for Liver
            Research, Kansai Medical University; Moriguchi, Osaka 570, Japan
            (Tel:06-992-1001(ex.2530), Fax:06-994-6099)
FEATURES             Location/Qualifiers
     source          1..483
```

```
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /clone="lambda hp44-1, lambda hp44-2"
          CDS       join(D28908.1:924. .1380,D28909.1:82. .118,D28910.1:50. .245,
                    D28911.1:139. .288,D28912.1:55. .226,D28913.1:115. .215,
                    D28914.1:121. .295,59. .105)
                    /codon_start=1
                    /product="hepatitis C-associated microtubular aggregate
                    protein p44"
                    /protein_id="BAA06043.1"
                    /db_xref="GI:1359443"
                    /translation="MAVTTRLTWLHEKILQNHFGGKRLSLLYKGSVHGFRNGVLLDRC
                    CNQGPTLTVIYSEDHIIGAYAEESYQEGKYASIILFALQDTKISEWKLGLCTPETLFC
                    CDVTKYNSPTNFQIDGRNRKVIMDLKTMENLGLAQNCTISIQDYEVFRCEDSLDERKI
                    KGVIELRKSLLSALRTYBPYGSLVQQIRILLGPIGAPKSSFFNSVRSVFQGHVTHQA
                    LVGTNTTGISEKYRTYSIRDGKDGKYLPFILCDSLGLSEKEGGLCRDDIFYILNGNIR
                    DRYQFNPMESIKLNHHDYIDSPSLKDRIHCVAFVFDASSIQYFSSQMIVKIKIQREL
                    VNAGVVHVALLTHVDSMDLITKGDLIEIERCEPVRSKLBEVQRKLGFALSDISVVSNY
                    SSEWELDPVKDVLILSALRRMLWAADDFLEDLPFEQIGNLREEIINCAQGKK"
          exon      59. .361
                    /number=9
          polyA_signal  344. .349
          polyA_site    361
ORIGIN
```

```
Query Match    14.7%;   Score 364;   DB 9;   Length 483;
Best Local Similarity 100.0%;   Pred. No. 1.2e-61;
Matches  364;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   2121 CCTCTGGTTGCCTTTCCTGAGATAATCCACTAAGAAATATTTTGTTCTCTTTTCTCAGGG 2180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCTCTGGTTGCCTTTCCTGAGATAATCCACTAAGAAATATTTTGTTCTCTTTTCTCAGGG 60

Qy   2181 AATCTAAGGAGGAGAAATTATCAACTGTGCACAAGGAAAAAATAGATATGTGAAAGGTTC 2240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AATCTAAGGAGGAGAAATTATCAACTGTGCACAAGGAAAAAATAGATATGTGAAAGGTTC 120

Qy   2241 ACGTAAAATTCCTCACATCACCAGAAGATTAAAATTCAGAAAGGAGAAAACACAGACCAAA 2300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACGTAAAATTCCTCACATCACCAGAAGATTAAAATTCAGAAAGGAGAAAACACAGACCAAA 180

Qy   2301 GAGAAGTATCTAAGACCAAAGGGATGTGTTTTATTAATGCTCTAGGATGAAGAAATGCATA 2360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAGAAGTATCTAAGACCAAAGGGATGTGTTTTATTAATGCTCTAGGATGAAGAAATGCATA 240

Qy   2361 GAACATTGTAGTACTTGTAAATAACTAGAAAATAACATGATTTAGTCATTAATTGTGAAAAA 2420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAACATTGTAGTACTTGTAAATAACTAGAAAATAACATGATTTAGTCATTAATTGTGAAAAA 300

Qy   2421 TAATAATAATTTTTCTTGGATTTATGTTCTGTATCTGTGAAAAAATAAATTTCTTATAAA 2480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TAATAATAATTTTTCTTGGATTTATGTTCTGTATCTGTGAAAAAATAAATTTCTTATAAA 360

Qy   2481 ACTC 2484
          ||||
Db    361 ACTC 364
```

```
RESULT 15
AX284296
LOCUS       AX284296                   580 bp     DNA        linear    PAT 20-NOV-2001
DEFINITION  Sequence 101 from Patent WO0179556.
ACCESSION   AX284296
VERSION     AX284296.1   GI:17044984
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Lillie,J., Brown,J.L., Bolt,A. and van Huffel,C.
TITLE       Novel genes, compositions and methods for the identification,
```

```
JOURNAL     assessment, prevention, and therapy of human cancers
            Patent: WO 0179556-A 101 25-OCT-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES    Location/Qualifiers
     source          1. .580
                    /organism="Homo sapiens"
                    /mol_type="unassigned DNA"
                    /db_xref="taxon:9606"
ORIGIN
```

```
Query Match    13.3%;   Score 331.2;   DB 6;   Length 580;
Best Local Similarity 96.4%;   Pred. No. 3.6e-55;
Matches  380;  Conservative  0;  Mismatches  10;  Indels  4;  Gaps  4;

Qy   1123 GGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTCGACATCTCCGGTGGTTAGCAATTATTC 1182
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    188 GGAAGTCCAAAGAAAACTTGGATTTGCTCTTCTCGACATCTCGGTGGTTAGCAATTATTC 247

Qy   1183 CTCTGAGTGGGAGCTGGAGCCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGAGACGAAT 1242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    248 CTCTGAGTGGGAGCTGGACCCTGTAAAGGATGTTCTAATTCTTTCTGCTCTGAGACGAAT 307

Qy   1243 GCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTTTGAGCAAATAGGTAGATGGTT 1302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    308 GCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCCTTTTGAGCAAATAGGTAGATGGTT 367

Qy   1303 TGGTGGTGT-GGAAGCTTGGAAGCCGGTCAGGTAGTTGGCTACTTTCGTCTTGGATCTATT 1361
          |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    368 TGGTGGTGTGGGGAAGCCTGGAAGCCGGTCAGGTAGTTGGCTACTTTCGTCTTGGATCTATT 427

Qy   1362 AAATACCTGGCAGCTCCTCGTCTTTTGTTTCTTTTCTCGCTTT 1421
          |||||||||||||||||||||||  |||||||||||||||||||
Db    428 AAATACCTGGCAGCTCCTCGTC-TTTTGTGGGTTGTTGCCCTG GAATAGTTCTGCTTTT 486

Qy   1422 TAACCCACTCCCTGGATGCATTTTCCCTCCTTGCATTTCCCTCCTTTTCCTGG-AGTTCA 1480
          ||||||||||||||||||||  |||||||||||||||||||  ||||||||||||||||||
Db    487 TAACCCACTCCCTGGATGCATTTTNCCTTCTGCATTTTCCTCCTTTTCCTGGAAGTTCA 546

Qy   1481 TACT-AGAGAATCTGCACTAATGTTTTTCCCTTT 1513
          |||| |||||||||||||||||||||||||||||
Db    547 TACTAAANAGAATCTGCACTAATGGTTTTCCCTTT 580
```

```
Search completed: April 25, 2004, 18:58:36
Job time : 6545.9 secs
```

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:00:48 ; Search time 2810.53 Seconds
(without alignments)
16458.288 Million cell updates/sec

Title: US-10-051-835-17
Perfect score: 1549
Sequence: 1 aacacatcagatatttcag.........ctctaaatctttaccatcaa 1549

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27511289 seqs, 1493109276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
```
EST:*
  1:  em_estba:*
  2:  em_esthum:*
  3:  em_estin:*
  4:  em_estmu:*
  5:  em_estov:*
  6:  em_estpl:*
  7:  em_estro:*
  8:  em_htc:*
  9:  gb_est1:*
 10:  gb_est2:*
 11:  gb_htc:*
 12:  gb_est3:*
 13:  gb_est4:*
 14:  gb_est5:*
 15:  em_estfun:*
 16:  em_estom:*
 17:  em_gss_hum:*
 18:  em_gss_inv:*
 19:  em_gss_pln:*
 20:  em_gss_vrt:*
 21:  em_gss_fun:*
 22:  em_gss_mam:*
 23:  em_gss_mus:*
 24:  em_gss_pro:*
 25:  em_gss_rod:*
 26:  em_gss_phg:*
 27:  em_gss_vrl:*
 28:  gb_gss1:*
 29:  gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	889	57.4	1200	13	BX447355	BX447355 BX447355
2	814.2	52.6	1062	12	BM471130	BM471130 AGENCOURT
3	617.2	39.8	972	10	BE888763	BE888763 601513677
4	614.4	39.7	616	9	AL603127	AL603127 DKFZp686D
5	601.6	38.8	601	14	CB134607	CB134607 K-EST0186
6	570.6	36.8	611	13	BX473904	BX473904 DKFZp686P
7	538	34.7	538	14	CB123846	CB123846 K-EST0172
8	489.6	34.3	960	13	BX457116	BX457116 BX457116
9	433.6	31.6	3953	11	AK052535	AK052535 Mus muscu
10	415.4	28.0	768	9	AL701060	AL701060 DKFZp686I
11	377.6	26.8	602	12	BG615970	BG615970 602643421
12	368.2	24.4	541	14	CD706751	CD706751 EST23278
13	358.6	23.8	664	10 c	AW532890	AW532890 UI-R-BU0-
14	351.8	23.2	572	9	AL700018	AL700018 DKFZp686O
15	335.6	22.7	553	13	BX473839	BX473839 DKFZp686H
16	330	21.7	545	10 c	BE683286	BE683286 182767 MA
17	326.6	21.3	624	14	CB105270	CB105270 K-EST0112
18	326.6	21.1	624	9	AL700011	AL700011 DKFZp686N
19	320.8	21.1	440	9	AL700902	AL700902 DKFZp686N
20	309	20.7	839	10 c	BF562499	BF562499 UI-R-BU0-
21	273	19.9	308	29 c	CC523935	CC523935 CH240_372
22	267.6	17.6	618	9	AI370412	AI370412 qv77d02.x
23	260.2	17.3	705	28 c	AZ383370	AZ383370 1M041D04
24	246.2	16.8	379	14 c	CF745483	CF745483 UI-M-GV0-
25	239.4	15.9	361	10 c	AW526795	AW526795 UI-R-BO1-
26	190.8	15.5	739	10	BE097316	BE097316 UI-R-BO1-
27	118.8	12.3	182	28 c	BZ206599	BZ206599 CH230-436
28	117	7.7	117	10 c	BF562500	BF562500 UI-R-BU0-
29	101	7.6	785	12	BG272761	BG272761 nah89d09.
30	93.2	6.5	220	28 c	BZ608978	BZ608978 WHADB05TR
31	90.6	6.0	786	10 c	AW260345	AW260345 um77b01.x
32	88.2	5.8	1811	14	CK018177	CK018177 AGENCOURT
33	84.2	5.7	408	29 c	CG753732	CG753732 P048-4-G0
34	77.8	5.4	295	14	CB770357	CB770357 AMGNMUC:N
35	74.8	5.0	1124	12 c	BG942089	BG942089 ax21a01.x
36	73.8	4.8	1566	13	BX436282	BX436282 BX436282
37	73.6	4.8	495	29 c	CG757757	CG757757 P053-1-D0
38	73.4	4.8	1531	29	BX217750	BX217750 Danio rer
39	73	4.7	187	29	CG748014	CG748014 P041-4-B0
40	71.6	4.7	342	14 c	CD701571	CD701571 EST18095
41	71.6	4.6	437	29	AY406818	AY406818 Homo sapi
42	71.6	4.6	453	9 c	AI087313	AI087313 oz45e10.x
43	71.6	4.6	467	10	BE675526	BE675526 7f10g07.x
44	71.6	4.6	499	9 c	AI457736	AI457736 tj49a01.x
45	71.6	4.6		9 c	AI827737	AI827737 wf11e07.x

ALIGNMENTS

```
RESULT 1
BX447355
LOCUS       BX447355     1200 bp    mRNA      linear    EST 22-MAY-2003
DEFINITION  BX447355 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP006YK08
            5-PRIME, mRNA sequence.
ACCESSION   BX447355
VERSION     BX447355.1  GI:31031827
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1200)
AUTHORS     Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 4292.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0AW13ZH08QP1&cluster=4292.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
```

FEATURES   Faraday Avenue Genoscope sequence ID : CS0AAW13ZH08QP1.
source   Location/Qualifiers
            1..1200
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0CAP006YK08"
            /tissue_type="THYMUS"
            /clone_lib="Homo sapiens THYMUS"
            /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
            with a NotI-oligo(dT) primer. Five prime end enriched,
            double-strand cDNA was digested with Not I and cloned into
            the Not I and EcoRV sites of the pCMVSPORT 6 vector.
            Library was not normalized."

ORIGIN

Query Match   57.4%; Score 889; DB 13; Length 1200;
Best Local Similarity 99.3%; Pred. No. 3.9e-149;
Matches 903; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy   68 AGACTACTGGTATCAGAATGGCAAAGCAACTAACCTTCCAGAAAATACAGATGATTGG 127
           ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   282 AGTCAGTTTGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGG 341

Qy   128 ACAAAAGAGGATGTAAATCAGTGGTTAGAAGTCATAAGATTGACCAAAAACACAGGGAA 187
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   342 ACAAAAGAGGATGTAAATCAGTGGTTAGAAGTCATAAGATTGACCAAAAACACAGGGAA 401

Qy   188 ATTTTGACTGAACAAGACGTGAATGGAGACGTCTTGAAGTGGTTAAAAAAAGAACATCTT 247
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   402 ATTTTGACTGAACAAGACGTGAATGGAGACGTCTTGAAGTGGTTAAAAAAAGAACATCTT 461

Qy   248 GTTGATATGGGCATCACACATGGACCAGCTATTCAAATGAAGAAGTATTCAAAGAATTG 307
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   462 GTTGATATGGGCATCACACATGGACCAGCTATTCAAATGAAGAACTATTCAAAGAATTG 521

Qy   308 CGGAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAAT 367
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   522 CGGAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAAT 581

Qy   368 GCTCCTAAAGACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACAA 427
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   582 GCTCCTAAAGACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACAA 641

Qy   428 AAGGGTAAAGAAGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGT 487
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   642 AAGGGTAAAGAAGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGT 701

Qy   488 TCTAAGTCACTAAAAGTTGAGCTCATAGAAGTAAAATAGATTATACAAAGGAAAGGCAA 547
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   702 TCTAAG CACTAAAAGTTGAGCTCATAGAAGTAAAATAGATTATACAAAGGAAAGGCAA 761

Qy   548 CCATCCATAGACCTGACATGTGTATCATCATCCATTGATGAATTCAGTAATCATATCGT 607
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   762 CCATCCATAGACCTGACATGTGTATCATCATCCATTGATGAATTCAGTAATCATATCGT 821

Qy   608 TACAAGTTGGATTTAGTCTCACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATA 667
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   822 TACAAGTTGGATTTAGTCTCACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATA 881

Qy   668 CATGAATTCAAAGCCTTCACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAA 727
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   882 CATGAATTCAAAGCCTTCACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAA 941

Qy   728 TTTAGCAATGAGGTTTTCCGATTTGCTTCAGCTTGTGATGAATTCACGTACCAATGCACT 787
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   942 TTTAGCAATGAGGTTTTCCGATTTGCTTCAGCTTGTGATGAATTCACGTACCAATGCACT 1001

Qy   788 ATTCATTTTGGAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAAT 847
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1002 ATTCATTTTGGAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAAT 1061

Qy   848 GATACCAAGGAAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTTGAAGAC 907
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1062 GATACCAAGGAAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTTGAAGAC 1121

Qy   908 CATCAAGTCCAACAAGCAAAGAAGAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTG 967
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1122 CATCAAGTCCAACAAGCAAAGAAGAGTGCATTCGAGAGCCAAGATTTGTGGAAG-TTTACTG 1180

Qy   968 CCAAATAGT 976
           |||||||||
Db   1181 CCAAATAGT 1189

RESULT 2
BM471130
LOCUS   BM471130   6463894 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5517666
DEFINITION AGENCOURT_6463894 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5517666
            5', mRNA sequence.
ACCESSION   BM471130
VERSION   BM471130.1 GI:18520172
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1062)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12177 row: d column: 19
            High quality sequence stop: 667.
FEATURES   Location/Qualifiers
source   1..1062
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5517666"
            /tissue_type="leiomyosarcoma"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_71"
            /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 2.1 kb."

ORIGIN

Query Match   52.6%; Score 814.2; DB 12; Length 1062;
Best Local Similarity 94.9%; Pred. No. 9.1e-136;
Matches 897; Conservative 0; Mismatches 38; Indels 10; Gaps 5;

Qy   187 AATTTTGACTGAACAAGACGTGAATGGAGACGTCTTGAAGTGGTTAAAAAAAGAACATCT 246
           | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AATTTTGACTGAACAAGACGTGAATGGAGACGTCTTGAAGTGGTTAAAAAAAGAACATCT 60

Qy   247 TGTTGATATGGGCATCACACATGGACCAGCTATTCAAATAGAACTATTCAAAGAATT 306
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   61 TGTTGATATGGGCATCACACATGGACCAGCTATTCAAATAGAACTATTCAAAGAATT 120

Qy   307 GCGGAAAACAGCCATTGAAGATTCAGATTCAGACATCTCAGAATGGGAAAGCCCAGTAAAAA 366
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   121 GCGGAAAACAGCCATTGAAGATTCAGATTCAGACATTCAGAATGGGAAAGCCCAGTAAAAA 180

Qy   367 TGCTCCTAAAGACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACA 426
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGCTCCTAAAGACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACA 240

```
                                        http://image.llnl.gov
                                        Plate: LLAM9737  row: j   column: 20
                                        High quality sequence stop: 637.
                         FEATURES                Location/Qualifiers
                              source            1..972
                                                /organism="Homo sapiens"
                                                /mol_type="mRNA"
                                                /db_xref="taxon:9606"
                                                /clone="IMAGE:3914995"
                                                /tissue_type="leiomyosarcoma"
                                                /lab_host="DH10B (phage-resistant)"
                                                /clone_lib="NIH_MGC_71"
                                                /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                                                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                                                Average insert size 2.1 kb. "

                         ORIGIN

                         Query Match      39.8%;   Score 617.2;  DB 10;  Length 972;
                         Best Local Similarity 89.3%;  Pred. No. 1.4e-100;
                         Matches 725;  Conservative   0;  Mismatches  73;   Indels  14;   Gaps   5;

Qy     427 AAAGGGTAAAGAAGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGG 486     Qy     447 ATATGGCTAATCC-GTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTT 505
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                 || |||||||||| |||||||||||||||||||||||||||| |||||||||||||||||
Db     241 AAAGGGTAAAGAAGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGG 300     Db       1 ATATGGCTAATCCGGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTT 60

Qy     487 TTCTAAGTCACTAAAAGTTGAGCTCATAGAGAGATAAAATAGATTATACAAAGGAAAGGCA 546     Qy     506 GAGCTCATAGAGAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA 565
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTCTAAGTCACTAAAAGTTGAGCTCACAGAGAGATAAAATAGATTATACAAAGGAAAGGCA 360     Db      61 GAGCTCATAGAGAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA 120

Qy     547 ACCATCCATAGACCTGACATGTGTATCATATCCATTCATTGATGAATTCAGTAATCCATATCG 606     Qy     566 TGTGTATCATATCCATTGATGAATTCAGTAATCCATTCATTGATGAATTCAGTAATTGGATTTTAGT 625
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ACCATCCATAGACCTGACATGTGTATCATATCCATTCATTGATGAATTCAGTAATCCATATCG 420     Db     121 TGTGTATCATATCCATTGATGAATTCAGTAATCCATTCATTGATGAATTCAGTAATTGGATTTTAGT 180

Qy     607 TTACAGCCTGAATTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGAT 666     Qy     626 CTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATCAATCTCATTGATCCGATCAAAGCCTTC 685
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TTACAGCCTGAATTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGAT 480     Db     181 CTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATCAATCTCATTGATCCGATCAAAGCCTTC 240

Qy     667 ACATGAATTCAAAGCCTTCACAAATACAGCAACAGCCACAGAGAGAGATGTCAAGATGAA 726     Qy     686 ACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGCAATGAAATTTAGCAATGAGAGTTTTC 745
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ACATGAATTCAAAGCCTTCACAAATACAGCAACAGCCACAGAGAGAGATGTCAAGATGAA 540     Db     241 ACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGCAATGAAATTTAGCAATGAGAGTTTTC 300

Qy     727 ATTTAGCAATGAAGTTTTCCGATTTGCTTCAGCTTGTTATTG-AAGTGGCACATTATTCCACAGTTC 786     Qy     746 CGATTTGCTTCAGCTTGTATGAAATTGTTGGCATCAAGTCACGTACCAAGTCACGTACCAAGTCAA 805
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATTTAGCAATGAAGTTTTCCGATTTGCTTCAGCTTGTTCAGCTTGTTCAGTTGAAAGTGGCACATTATTCCACAATGGCAC 600     Db     301 CGATTTGCTTCAGCTTGTATGAAATTGTTGGCATCAAGTCACGTACCAAGTCACGTACCAAGTCAA 360

Qy     787 TATTCATTTGGAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAGTCACCAA 846     Qy     806 GACAAACCCATGGGAAAATTGTTGGCATCAAGTCACCAAGTCATCAAGTCATCCAACAAGCAGCCCTC 865
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TATTCATTTGGAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAGTCACCAA 660     Db     361 GACAAACCCATGGGAAAATTGTTGGCATCAAGTCACCAAGTCATCAAGTCATCCAACAAGCAGCCCTC 420

Qy     847 TGATACCCAAGGAAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTTGAAGA 906     Qy     866 ATTAACCATTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCAACAAGCA 925
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TGATACCCAAGGAAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTTGAAGA 720     Db     421 ATTAACCATTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCAACAAGCA 480

Qy     907 CCATCAAGTCCAACAAGCAAAGAAGTGCATTGCTTCGAGAGCCAAGATTTGTGGGAAGTTTTACT 966     Qy     926 AAGAAGTGCATTGCTTCGAGAGCCAAGATTTGTGGAAGTTTACTGGAAGTTTACTGGAAGTTTACTACTCTATCT 985
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CCATCAAGTCCAACAAGCAAAGAAGTGCATTGCTTCGAGAGCCAAGATTTGTGGGAAGTTTACT 780     Db     481 AAGAAGTGCATTGCTTCGAGAGCCAAGATTTGTGGAAGTTTACTGGAAGTTTACTGGAAGTTTACTACTCTATCT 540

Qy     967 GCCAAATAGTACTCTATCTGA-CAGATTTGTATTG-AAGTGGCACATTATTCCACAGTTC 1024    Qy     986 GACAGATTTGTT-ATTGAAGTGGACATTATTCCAGTTCTCTGAATGCCAATATGATTA 1044
           ||||||||||||||||||||| |||||||||||| ||||||||||||||||| |||||||                 ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GCCCAATAGGACTCTATCTGACCAGATTTGTGTATGTGTATTGTTATTGAAATGCCAAAATTTCCCACCGTTC 840    Db     541 GACAGATTTGTTGTTGCGGAGAATGGAACATGTCCACGGACACTTCCGCAAGACTAAAG 600

Qy     1025 T--CTGAATGCCAATATGATTAC-TTCCAGATTAAAATGCAAAT-----ACACAACA 1076    Qy     1045 CTTCCGATTAAAATGCAAAATGCAAAATTACACAACAACAATATGG-----AACAAAGTAAAAA 1099
            |  ||||||||||||||||||| ||||||||||||||||| |    ||||||||                 ||||||||||||||||||||||||||||||||||||||||||           ||||||||||
Db     841 TTCTGAAAGGCCAATATGATTACTTCCGGATTAAAATGCAAATGCCAAAATTTCCCCACCAA 900    Db     601 TTTCCGATTAAAATGCAAACTTACAACTTACAACAACAAATATGGGGAACCAACAAGTAAAACACA 660

Qy     1077 AAATATGGGAACAAAGTAAAAAAATTCTCACTATTGTGTGCGAGATG 1121    Qy     1100 TTCTCACTATTGTGTGCGGAG--ATGGACCAGCATTACGAAAA-----ATAA 1152
            |||| ||||||||||||||||||  |||||||||||||||||||                |||||||||||||||||||||  ||||||||||||||||||||||||||     ||||
Db     901 TTATGGGAACCAAGTAAAAAAATTCTCCCTTATTTGGGCAAG 945    Db     661 TTCTCACGTATTGTTGTGCGGAGAATGGAACGTCCTACGGACACTTCCGCAAGACTAAAG 720

                                                                             Qy     1153 AGTTGATTCAGAGCATTAAAGCAGGATTTAAAACACTGCCAGAGTCCAGAAAAGCAGC 1212
                                                                                        |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
                                                                             Db     721 GTGGACTTCCCGAGCCTTAACACCTTGGGCAATCCCGATTCTTAAACACTTGGGCAATCCCGAAACAGACG 780
```

RESULT 3
BE888763
LOCUS   BE888763   972 bp   mRNA   linear   EST 20-OCT-2000
DEFINITION   6015l3677F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3914995 5',
               mRNA sequence.
ACCESSION   BE888763
VERSION   BE888763.1   GI:10345391
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 972)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:

```
Qy    1213  AGAGAAAAATTCAGAGCAAAACAAATAAAA 1244
            ||||| |||||||||||||||| |||||||
Db     781  GACAAAAATCTCCGCAGAGAAATTAAACAAGA 812
```

```
RESULT 4
AL603127
LOCUS        AL603127              616 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION   DKFZp686D092_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
             DKFZp686D092 5', mRNA sequence.
ACCESSION    AL603127
VERSION      AL603127.1  GI:1516633
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 616)
  AUTHORS    Wambutt,R., Heubner,D., Mewes,W., Weil,B. and Wiemann,S.
  TITLE      EST (Wambutt.R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: MIPS
             MIPS
             Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
             This is the 5' sequence of the clone insert.
             Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
             Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
             sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
             consortium of the German Genome Project.
             No s1 sequence available.
             This clone (DKFZp686D092) is available at the RZPD in Berlin,
             Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
             Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..616
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686D092"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN
Query Match     39.7%;  Score 614.4;  DB 9;  Length 616;
Best Local Similarity   99.8%;  Pred. No. 5e-100;
Matches   615;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy    140  GTAAATCAGTGGTTAGAAAGTCATAGAGATTGACCAAAAACACAGGGAAATTTTGACTGAA 199
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GTAAATCCGTGGTTAGAAAGTCATAGAGATTGACCAAAAACACAGGGAAATTTTGACTGAA 60
```

```
Qy    200  CAAGACGTGAATGGAGCAGTCTTGAGGTGGTTAAAAAAGAACATCTTGTTGATATGGGC 259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  CAAGACGTGAATGGAGCAGTCTTGAGGTGGTTAAAAAAGAACATCTTGTTGATATGGGC 120
```

```
Qy    260  ATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAGCC 319
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  ATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAGCC 180
```

```
Qy    320  ATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAGAC 379
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  ATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAGAC 240
```

```
Qy    380  CAAACTGTGTCCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAGGGTAAAGAG 439
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  CAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACAAAAAGGGTAAAGAG 300
```

```
Qy    440  AACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCAAGTCACTA 499
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  AACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCAAGTCACTA 360
```

```
Qy    500  AAAGTTGAGCTCATAGAAGAGAAAATAGATTATACAAGGAAAGGCAACCATCCATAGAC 559
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    361  AAAGTTGAGCTCATAGAGATAAAAATGATTATACAACGGAAAGGCAACCATCCATAGAC 420
```

```
Qy    560  CTGACATGTGTATCATATCCATTTGATGAATTCAGTAGTTCCATATCGTTACAAGTTGGAT 619
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CTGACATGTGTATCATATCCATTTGATGAATTCAGTAGTTCCATATCGTTACAAGTTGGAT 480
```

```
Qy    620  TTTAGTCTCACAGCTGAAACAGAGACCAGGCAATCTCATTGATCCGATACATGAATTCAAA 679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  TTTAGTCTCACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCAAA 540
```

```
Qy    680  GCCTTCACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGCAATGAATTTAGCAATGAG 739
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  GCCTTCACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGCAATGAATTTAGCAATGAG 600
```

```
Qy    740  GTTTTCCGATTTGCTT 755
           ||||||||||||||||
Db    601  GTTTTCCGATTTGCTT 616
```

```
RESULT 5
CB134607
LOCUS        CB134607              601 bp    mRNA    linear   EST 29-JAN-2003
DEFINITION   K-EST0186079 L14ChoiCK0 Homo sapiens cDNA clone L14ChoiCK0-13-G04
             5', mRNA sequence.
ACCESSION    CB134607
VERSION      CB134607.1  GI:28100932
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 601)
  AUTHORS    Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
             Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
             Kim,Y.S.
  TITLE      21C Frontier Korean EST Project 2001
  JOURNAL    Unpublished (2002)
COMMENT      Contact: Kim YS
             Genome Research Center
             Korea Research Institute of Bioscience & Biotechnology
             52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
             Tel: +82-42-860-4470
             Fax: +82-42-860-4409
             Email: yongsung@mail.kribb.re.kr
             Plate: 13  row: G  column: 04
             High quality sequence stop: 601.
FEATURES             Location/Qualifiers
     source          1..601
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L14ChoiCK0-13-G04"
                     /sex="M"
                     /cell_line="Choi-CK"
                     /lab_host="Top10F'"
                     /clone_lib="L14ChoiCK0"
                     /note="Organ: Liver; Vector: pTT3-Pac; Site_1: EcoRI;
                     Site_2: NotI; The library was contributed by the Soares
                     laboratory and it was constructed as described by Bonaldo,
                     M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
                     6(9): 791-806. RNA was prepared from harvested cell
                     culture."
ORIGIN

Query Match     38.8%;  Score 601;  DB 14;  Length 601;
Best Local Similarity   100.0%;  Pred. No. 1.2e-97;
Matches   601;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    87  TGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGGATGTAAATC 146
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 6

```
Db   1 TGGCAAAGCAACTTAACCTTCCAGAAATACAGATGATTGGACAAAGAGGATGTAAATC 60
Qy 147 AGTGGTTAGAAAGTCATAAGATCATAAGATGGACCAAAGAAATTTGACTGAACAAGACG 206
       |||||||||||||||||||||         |||||||||||||||||||||||||||||
Db  61 AGTGGTTAGAAAGTCATAAGAT     TGACCAAAGAAATTTGACTGAACAAGACG 120

Qy 207 TGAATGGAGCAGTCTTGAAGTGTTAAAAAAAGAACATCTTGTTGATATGGGCATCACAC 266
       |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db 121 TGAATGGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTTGTTGATATGGGCATCACAC 180

Qy 267 ATGGACCAGCTATTCAAATAGAAGAACATTCAAAGAATTGCGGAAAACAGCCATTGAAG 326
       |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db 181 ATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAGCCATTGAAG 240

Qy 327 ATTCGATTCAGACATCTAAGAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAGACCAAACTG 386
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 ATTCGATTCAGACATCTAAGAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAGACCAAACTG 300

Qy 387 TGTCTCAAAAGGAACGTAGAGAGAAACTTCAAAGAGGGTAAAGCAAAACAAAAGAAACCCAG 446
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 TGTCTCAAAAGGAACGTAGAGAGAAACTTCAAAGAGGGTAAAGCAAAACAAAAGAAACCCAG 360

Qy 447 ATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAGGTTCTAAGTCACTAAAAGTTG 506
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 ATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAGGTTCTAAGTCACTAAAAGTTG 420

Qy 507 AGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATCCATAGACCTGACAT 566
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 AGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATCCATAGACCTGACAT 480

Qy 567 GTGTATCATATCCATTGATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTTAGTC 626
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 GTGTATCATATCCATTGATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTTAGTC 540

Qy 627 TACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGATCTCATTGAATTCAAAGCCTTCA 686
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGATCTCATTGAATTCAAAGCCTTCA 600

Qy 687 C 687
Db 601 C 601
```

ORIGIN

```
Query Match            36.8%;   Score 570.6;  DB 13;  Length 611;
Best Local Similarity  99.3%;   Pred. No. 3.3e-92;
Matches  573;  Conservative  0;  Mismatches  4;  Indels  0;  Gaps  0;

Qy   1 AACACATCAGATATTTCAGCACTAAAAGAAGATGGTTTCCCACATATGTAAAAGAA 60
       ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  35 AACACATCAGATATTTTCAGCACTAAAAGAAGATGGTTTCCCACATATGTAAAAGAA 94

Qy  61 ATTTGCAAGACTACTGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGA 120
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  95 ATTTGCAAGACTACTGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGA 154

Qy 121 TGATTGGACAAAGAGGATGTAAATCAGTGGTTAGAAGTGATTCATAAGAGTTGACCAAAACA 180
       |||||||||||||||||||||||||||||||||||  |||||||||||   |||||||||||
Db 155 TGATTGGACAAAGAGGATGTAAATCAGTGGTTAGAAAGTGATTCATAAAGTGATTGACCAAAACA 214

Qy 181 CAGGGAAATTTGACTGAACAAGACGTGAATGGCAGTCTTGAAGTGGTTAAAAAAAGA 240
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 215 CAGGGAAATTTGACTGAACAAGACGTGAATGGCAGTCTTGAAGTGGTTAAAAAAAGA 274

Qy 241 ACATCTTGTTGATATGGGCATCACACATGAAGATTCAAATAGAAGAACTATTCAA 300
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 275 ACATCTTGTTGATATGGGCATCACACATGAAGATTCAAATAGAAGAACTATTCAA 334

Qy 301 AGAATTGCGGAAAAACAGCCATTGAAGATTCAAGATTCAGAGATGGGAAAGCCCAG 360
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 335 AGAATTGCGGAAAAACAGCCATTGAAGATTCAAGATTCAGAGATGGGAAAGCCCAG 394

Qy 361 TAAAAATGCTCCTAAAGACCAAACTGTGTCTCAAAAGGACGTAGAGAGAAACTTCAAAGCA 420
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 395 TAAAAATGCTCCTAAAGACCAAACTGTGTCTCAAAAGGACGTAGAGAGAAACTTCAAAGCA 454

Qy 421 AAAACAAAAGGGTAAAGACGTAAAGAGAAACCCAGATATGGCTAATCCGTCCAATGAGCTACAACTGC 480
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 455 AAAACAAAAGGGTAAAGACGTAAAGAGAAACCCAGATATGGCTAATCCGTCCAATGAGCTACAACTGC 514

Qy 481 TAAAGGTTCTAAGTCACTAAAAGTTGAGCTCATAGAAGAAATAGATTATATACAAAGGA 540
       ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db 515 TAAAGGTTCTAAGTCACTAAAAGTCGAGCTCATAGAAGAAATAGATTATATACAAAGGA 574

Qy 541 AAGGCAACCATCCATAGACCTGACATGTGTATCATAT 577
       |||||||||||||||||||||||||||||||||||||
Db 575 AAGGCAACCATCCATAGACCTGACATGTGTATCATAT 611
```

```
DEFINITION   BX457116 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP006YK08
             5-PRIME, mRNA sequence.
ACCESSION    BX457116
VERSION      BX457116.1  GI:31030860
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 960)
AUTHORS      Li,W.B., Gruber,C., Jessee,J. and Polaves,D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
             Genoscope - Centre National de Sequencage
             BP 191 91006 EVRY cedex - France
             Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
             Library was constructed by Life Technologies, a division of
             Invitrogen. This sequence belongs to sequence cluster 4292.r For
             more information about this cluster, see
             http://www.genoscope.cns.fr/
             cgi-bin/cluster.cgi?seq=CS0CAP006BF04QP1&cluster=4292.r. Contact :
             Feng Liang Email : fliang@lifetech.com URL :
             http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
             Faraday Avenue Genoscope sequence ID : CS0CAP006BF04QP1.
FEATURES             Location/Qualifiers
     source          1..960
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP006YK08"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

Query Match          34.3%;   Score 531;  DB 13;  Length 960;
Best Local Similarity   93.2%;   Pred. No. 3.5e-85;
Matches   573;  Conservative   22;  Mismatches   16;  Indels   4;  Gaps   4;

Qy      68 AGACTACTCGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGG 127
           ||::   :|||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 AGDSAGWWTGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGG 358

Qy     128 ACAAAAGAGGATGTAAATCAGTGGTTAGAAAGTTCATAAGATTGACCAAAAACACAGGGAA 187
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     359 ACAAAAGAGGATGTAAATCAGKGGTTAGAAAGTTCATAAGATTGACCAAAAACACAGGGAA 418

Qy     188 ATTTTGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAA-AAGAACATCT 246
           |||||||||||||||||||||||||||||||||||||  :|||||:||||||
Db     419 ATTTTGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAARRAGAACATCT 478

Qy     247 TGTTGATATGGGCATCACCAT-GGACCAGCTATTCAAATAGAAGAACTATTCAAAGAAT 305
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     479 TGTTGATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAAT 538

Qy     306 TGCGGAAAACAGCCATTGAAGATTCGATTCAGACATCAAGATGGGAAAGCTAGAAACTT 365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     539 TGCGGAAAACAGCSATTGAAGATTCGATTCAGACATCAAGATGGGAAAGCCCAGTAAAA 598

Qy     366 ATGCTCCTAAAGACC-AAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAA 424
           |||||||||||||||  :||||||||||||||||||||||||||||||||||||||||
Db     599 ATGCTCCTAAAGACCAAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAA 658

Qy     425 CAAAAGGGTAAAGAGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAA 484
           |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db     659 CAAAAGGGKRAAGAGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAA 718

Qy     485 GGTTCTAAGTCACTAAAAGTTGAGCTCATAGAGAATAAAATAGATTATACAAAGGAAAGG 544
```

```
TITLE        21C Frontier Korean EST Project 2001
JOURNAL      Unpublished (2002)
COMMENT      Contact: Kim YS
             Genome Research Center
             Korea Research Institute of Bioscience & Biotechnology
             52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
             Tel: +82-42-860-4470
             Fax: +82-42-860-4409
             Email: yongsung@mail.kribb.re.kr
             Plate: 31 row: D column: 07
             High quality sequence stop: 538.
FEATURES             Location/Qualifiers
     source          1..538
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L10ChoCK0-31-D07"
                     /sex="M"
                     /cell_line="Cho-CK"
                     /lab_host="Top10F'"
                     /clone_lib="L10ChoCK0"
                     /note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI;
                     Site_2: NotI; The library was contributed by the Soares
                     laboratory and it was constructed as described by Bonaldo,
                     M.F., Lennon, G. and Soares, M.B. (1996). Genome Research
                     6(9): 791-806. RNA was prepared from harvested cell
                     culture."
ORIGIN

Query Match          34.7%;   Score 538;  DB 14;  Length 538;
Best Local Similarity   100.0%;   Pred. No. 2.3e-86;
Matches   538;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     165 AGATTGACCAAAAACACAGGGAAATTTGATGAACAAGACGTGAATGGAGCAGTCTTGA 224
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGATTGACCAAAAACACAGGGAAATTTGATGAACAAGACGTGAATGGAGCAGTCTTGA 60

Qy     225 AGTGGTTAAAAAAGAACATCTTGTTGATATGGGACCATCACACATGGACCAGCTATTCAAA 284
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AGTGGTTAAAAAAGAACATCTTGTTGATATGGGACCATCACACATGGACCAGCTATTCAAA 120

Qy     285 TAGAAGAACTATTCAAAGAATTGCGGAAAACAGCCATTCGATTCAGACATCTA 344
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TAGAAGAACTATTCAAAGAATTGCGGAAAACAGCCATTCGATTCAGACATCTA 180

Qy     345 AGATGGGAAAGCCCAGTAAAAATGCTCCTAAAGACCAAACTGTGTCTCAAAAGGAACGTA 404
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGATGGGAAAGCCCAGTAAAAATGCTCCTAAAGACCAAACTGTGTCTCAAAAGGAACGTA 240

Qy     405 GAGAAACTTCAAAGCAAAAACAAAAGGGTAAAGAGAACCCAGATATGGCTAATCCGTCTG 464
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GAGAAACTTCAAAGCAAAAACAAAAGGGTAAAGAGAACCCAGATATGGCTAATCCGTCTG 300

Qy     465 CAATGAGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTTGAGCTCATTAGAAGATAAAA 524
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAATGAGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTTGAGCTCATTAGAAGATAAAA 360

Qy     525 TAGATTATACAAAGGAAAGGCAACCAACATCCATAGACCTGACATGTGTATCATATCCATTTG 584
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TAGATTATACAAAGGAAAGGCAACCAACATCCATAGACCTGACATGTGTATCATATCCATTTG 420

Qy     585 ATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTTAGTCTACAGCCTGAAACAGGAC 644
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTTAGTCTACAGCCTGAAACAGGAC 480

Qy     645 CAGGCAATCTCATTGATCCGATACATGAATTCAAAGCCTTCACAAATACAGCAACAGC 702
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CAGGCAATCTCATTGATCCGATACATGAATTCAAAGCCTTCACAAATACAGCAACAGC 538
```

```
RESULT 8
BX457116
LOCUS       BX457116               960 bp    mRNA      linear     EST 22-MAY-2003
```

```
Db    719 GGDWCTRAGTCACTAAAAGTTGAGCTCATTAGAAGAGATAAAATAGATTATDCAAAGGAAAGG  778
          ||::||||||    |||||||||||||||||||||||||:||||||||
Qy    545 CAACCATCATAGACCTGACATGTATCATATCCATTTGATGAATTCAGTAGTAATCCATAT  604
          ||||||||||||::|||||||:||:|||:||||||||||| ||:|||||||||||||||||
Db    779 CAACCATCCAWAGACCTGACATGTGTWTCATAWCCATTTGAGAAWTCAGTAATCCATAT  838

Qy    605 CGTTACAAGTTGGATTTTAGTCHACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCG  664
          ||||||||||||||||||||:||:|||:|||||||||:|||:|||||||||||:|:||
Db    839 CGTTACAAGDTGGGATTTWGTCTAMAG--CTGAAACAGGGMCAGGCMAFCTCATTGGTMCG  897

Qy    665 ATACATGAATTCAAA  679
          :||||||||||||||
Db    898 RAAAAAAAAAAA  912
```

RESULT 9

AK052535

LOCUS   AK052535   3953 bp   mRNA   linear   HTC 20-SEP-2003
DEFINITION   Mus musculus 13 days embryo lung cDNA, RIKEN full-length enriched
             library, clone:D43004G20 product:weakly similar to CDNA FLJ20073
             FIS, CLONE COLO2320 [Homo sapiens], full insert sequence.
ACCESSION   AK052535
VERSION   AK052535.1   GI:26095242
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
             Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsuni,T., Tashiro,T., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
             FANTOM Consortium.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6   (bases 1 to 3953)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
```

---

```
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shiragata,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,T., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M. and Hayashizaki,Y.
  TITLE      Direct Submission
  JOURNAL    Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
COMMENT      cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.go.jp/
             URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3953
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:D430044G20"
                     /db_xref="MGI:2422069"
                     /db_xref="taxon:10090"
                     /clone="D430044G20"
                     /tissue_type="lung"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="13 days embryo"
     misc_feature    328..>3952
                     /note="putative
                     weakly similar to CDNA FLJ20073 FIS, CLONE COLO2320 [Homo
                     sapiens] (SPTR|Q9NXS8, evidence: FASTY, 50.4%ID,
                     71.7%length, match=1341)"
ORIGIN
     Query Match             31.6%;   Score 489.6; DB 11;  Length 3953;
     Best Local Similarity   61.9%;   Pred. No. 6.3e-78;
     Matches  962;  Conservative    0;  Mismatches  549;  Indels   43;  Gaps   10;

Qy     15 TTTCAGCACTAAAAGAGAGATGGTTTTCCCCACATATATGTAAAAGAGAAATTGCAAGACTAC   74
          ||| ||||| ||||  ||||||  |||| |||||||||||||||||  |||| |||
Db    257 TTTTTTCCTTTTAAACGATTTTTTATGTAGAAGAAATTTGAAAATTGGGGAATAAATTT  316

Qy     75 TGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAAATACAGATGATTGGACAAAAG  134
          |||||||  |||||||||||| |||||| ||||||  |||||||||| |||||||||||||
Db    317 GAAGCTTGAACATGAGTGGACAAGTAACTCAACCAAAATTGATCAAAGACTGGACCAAAG  376

Qy    135 AGGATGTAAATCAGTGGTT--AGAAAGTCATAAGATTGACCAAAAACACAGGGAAATTT  191
          ||||||||| ||||||| |  ||||  |||| ||||  |||||||||| |||||||||
Db    377 AACATGTAAGAAAATGGGTTACTGAGCATCTTAAACAATTGTTGAGAAATATGCTCAAATCT  436

Qy    192 TGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTG  251
          ||||| ||||| |  |||| |||||  |||| ||||  |||| |||| |||||||||||
Db    437 TGTTCAAGAAGTAACAGGGATGGTCCTGCAGGAATTAACTGAGAAGATCTTAGAG  496

Qy    252 ATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGA  311
          |||||||||| | ||||| |||||||| ||||| |||| |||| ||| |||||||||||||
Db    497 AAATGGGCTTGCCACGGGGTCCAGCACTTCTGATAAAAACGCATGTATAAACAAA---CTAA  553

Qy    312 AAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGAAAGCCCAGTAAAAATGCTC  371
          ||||| |||||||||| ||||||  | ||| |||| ||||  |||||||||||||||||||
Db    554 TTAGTTCCCCTGAAAGTCACAATCAAGAATTCCAGAGAATTAAATGATAAAAAACTTCAA  613

Qy    372 CTAAAGACCAAACTGTGTCTCAAAGAAGACGTTAGAGAAACTTGAGAAAAACACAAAAGG  431
          |||||||||| | ||||| |||||| ||||||||| |||| |||| ||||| |||||||||
Db    614 CAAAAGAACAGCAAACAAAGACCAAAAATGAGGAAGAAAATGCAGTTTCATCCAACAGTG  673
```

Qy   1496 GAACAGAAAACCACCACCAAATGAGACGATTCTACTCTAAATCTTTACCATCAA 1549
            |||||||||||||||||||||||||||| |||||||||
Db   1736 GAAAAACTACCATAGCA---GAGAAAATTCTACTCTGAAACTTTATGAGCAA 1786

RESULT 10
AL701060
LOCUS   AL701060   762 bp   mRNA   linear   EST 04-SEP-2003
DEFINITION DKFZp686I14120_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686I14120 5', mRNA sequence.
ACCESSION   AL701060
VERSION   AL701060.1 GI:19621593
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 762)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and
            Wiemann,S.
  TITLE   EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and
            Wiemann,S.)
  JOURNAL   Unpublished (2001)
COMMENT   Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. No s1 sequence
            available.
            This clone (DKFZp686I14120) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES   Location/Qualifiers
     source   1..762
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686I14120"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

Query Match   28.0%; Score 433.6; DB 9; Length 762;
Best Local Similarity   75.4%; Pred. No. 9.3e-68;
Matches 573; Conservative   0; Mismatches 169; Indels 18; Gaps 2;

Qy   687 CAATACAGCAACAGCCACCAGAGAGGATGTCAAGATGAAATTAGACAATGAGGTTTTCC 746
           || ||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   1 CAACACAGAAAACAGCCACCACCCGAGTGGACATTAAGATGAAATTCAGCAATGAGTCTTCC 60

Qy   747 GATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAAAG 806
           |||||| ||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   61 GATTTGCATCAGCTTGTATGAATTCACGCACCAATGGCACCATCCATTTTGGAGTCAAGG 120

Qy   807 ACAAACCCATGGAGAAAATTGTTGGCATCAAAGTCACCAAGTACCAAGGAGAGCCCTCA 866
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   121 ACAAACCCATGGAGAAAATTGTTGGTGTGAAATATCACCAGT-----AAGGCTGCCTTCA 174

Qy   867 TTAACCATTTCAATCTGATGATAAACAGTATTTGAAGACCATCAAGTCCAACAAGCAA 926
           |||||||||||||||||||| ||||||||||||||||| |||||||||||||||||||
Db   175 TTGACCACTTCAATGTAATGATCAATAAAAGTATTTTGAAGAAAGTGAGATCAATGAAACA 234

Qy   927 AGAAGTGCATTCGAGAGCCAAGATTTGTTGGAAGTTTACTGCCAAATAGTACTTCTATCTG 986
           ||||||| |||||||||||||||||||||||||||| |||||||||||||| |||||
Db   235 AGAAGTGTATTCGGGAGCCAAGGTTTGTTGGGAAGTCCTTCTGCAGAACAATAACACCATCTG 294

Qy   987 ACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAATATGATTACT 1046

```
sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T).  Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."

ORIGIN

Query Match      26.8%;   Score 415.4;   DB 12;   Length 768;
Best Local Similarity   92.8%;   Pred. No. 1.7e-64;
Matches 491;   Conservative   0;  Mismatches  31;   Indels   7;  Gaps   5;

Qy    68 AGACTACTGGGTATCAGAAGGCGAAGCAACTTCCAGAAATACAGATGATTGG 127
            |||||| ||||||||||||||  ||||||||  |||||||||||||||||||||||
Db   240 AGTCAGTTGGTATCAGAATGCAAGCAACTAACCTTCCAGAAATACAGATGATGATTGG 299

Qy   128 ACAAAAGAGGATGTAAATCAGTGGTTAGAAGTCATAAGGTCATAAGAAAACACAGGGAA 187
            ||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||
Db   300 ACAAAAGAGGATGTAAATCAGTGGTTAGAAGTCATAAGATTGACCAAAAACACAGGGAA 359

Qy   188 ATTTTGACTGAACAAGACGTGAATGGAGCAGTCTTGGAAGTGGTTAAAGAAGAACATCTT 247
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   360 ATTTTGACTGAACAAGACGTGAATGGAGCAGTCTTGGAAGTGGTTAAAAAGAAGAACATCTT 419

Qy   248 GTTGATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTG 307
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   420 GTTGATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTG 479

Qy   308 CGGAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAGCCCAGTAAAAAT 367
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   480 CGGAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAGCCCAGTAAAAAT 539

Qy   368 GCTCCTAAAGACCAAACTTCAAAAAGAAACTTCAAAAGCAAAACAA 427
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db   540 GCTCCTAAAGACCAAACTTCAAAAAGAAACTTCAAAAGCAAAACAA 599

Qy   428 AAGGGT-AAAGAGAACCCAGATAT-GGCTAATCCGTCT-GCAATGAGTAC--AACTGCTA 482
            ||||||  |||||||||||||||| |||||||||||||  |||||||||||  ||||||||
Db   600 AAGGGTCACAGAGAACCCAGATATGGGCTAATCCGTCTGCCAATGAGTTCCAACTGCCTA 659

Qy   483 AAGGTTCTAAGTCACTA--AAAGTTGAGCTCATAGAGAATAAAATGATTATACAAAGGA 540
            ||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Db   660 AAGGTTCTAAGTCACTAACAGTTGAGGCTCATAGAGAAGATCAATGATTATACACAGGA 719

Qy   541 AAGGCAACCATCCATAGACCTGACATGTGTATCATATCCATTTGATGAA 589
            ||||||||||||||||||||| |||  |||   |||||||||||||||||
Db   720 AAGGCAACCATCCATAGACCCTGCCACTGGGTTCTATTCCCTTTGATGGA 768


RESULT 12
CD706751
LOCUS       CD706751        602 bp    mRNA    linear   EST 25-JUN-2003
DEFINITION  EST23278 human nasopharynx Homo sapiens cDNA, mRNA sequence.
ACCESSION   CD706751
VERSION     CD706751.1  GI:32237381
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 602)
AUTHORS     Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and
            Zeng,Y.-X.
TITLE       Transcriptional Gene Expression Profile of Human Nasopharynx
JOURNAL     Unpublished (2003)
COMMENT     Contact: Yixin Zeng
            Cancer Center
            Sun Yat-sen University
            651 DongFeng Road East, GuangZhou 510060, China
            Tel: 86-1380-9770-743
            Fax: 86-20-8775-4506
            Email: yxzeng@zsums.edu.cn.
```

```
Db   295 |||||||||||| |||||| |||||||||||||  |||| |||||||| 354
Qy  1047 TCCAGATTAAAATGCAAAATTACACACAACAAAATATGGGAACAAAGTAAAAAAATTCTCAC 1106
Db   355 TCTACATTCAGATGCAAATTTGTAAAGATAGATAGAACAAAACCAAAATCTTTCAC 414

Qy  1107 TATTTGTGCGGAGATGGGACCAGCTCTAAGGACATTACGAAAAATA---------AAG 1154
            ||||||| |||||||  |||||  ||||| |||||  |||    |||        |||
Db   415 TGTTTGTAAGAAGGGGCTAGCTTCCTAGGGATATCCTGGCCAATTCCAAGCAACGGGGATG 474

Qy  1155 TTGATTTCAGAGCATTTAAAGCAGATTTTAAACACTGCGAGTCCAGAAAGCAGCAG 1214
            ||||||||| |||||| |||||||||||| ||||||| ||||| |||||||||| |||
Db   475 TAGATTTCAAGGCATTTTTACAAATTTAAGTCACTGTAGCATCTAGAAAGAGAGGCTG 534

Qy  1215 AAGAAAATTCAGAGCAAAAACAAATAAAAAAAGAGAGGGACCAAAGTTGGTTAAAT 1274
            |||||| |||||| |||| |||||||||||||||||| ||||||| ||||||||||
Db   535 AAGAAGAGTATGGAATGCAATGAAGGCAATGAGAGGAGAGTGAAGGACTAAAGTCGGTTAAAC 594

Qy  1275 TATTGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTAA 1334
            | ||||| ||||| ||||||||  |||||||||||||||||  ||  ||||||||||||||
Db   595 TTCTCATAGGAAAACCGAGACTCACTGGATAATTCATACTATGACTGGTACATTCTTGTAA 654

Qy  1335 CAAATAAATGCCACCCAGATCAAACAAAACAACTTAGATTTCCTGAAGGAAATTAAATGGT 1394
            ||||||||||||| ||  | |||||| ||| ||||||||||   |||| ||||||||||||
Db   655 CAAATAAATGCCATCCAAACCAAACCAAATAAAGCACTTAGATTTTTAAAGAAAATTAAATGGT 714

Qy  1395 TTGCTGTGTTGGAGTTTGATCCTGAGTCTAACATCAATGG 1434
            ||||||||||||||||||||||||| ||||  |||||||||
Db   715 TTGCTGTGTTGGAGTTTGATCCTGAATCTATGATCAATGG 754


RESULT 11
BG615970
LOCUS       BG615970        768 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602643421F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4774373 5',
            mRNA sequence.
ACCESSION   BG615970
VERSION     BG615970.1  GI:13667354
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 768)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1647 row: j column: 06
            High quality sequence stop: 671.
FEATURES             Location/Qualifiers
     source          1..768
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4774373"
                     /tissue_type="embryonal carcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_61"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggcgcgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
```

```
FEATURES            Location/Qualifiers
     source         1..602
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /tissue_type="normal nasopharynx"
                    /clone_lib="human nasopharynx"
                    /note="ESTs generated from a normal nasopharynx cDNA
                    library from southern Chinese"

ORIGIN

Query Match           24.4%;    Score 377.6;  DB 14;  Length 602;
Best Local Similarity 97.7%;    Pred. No. 9.9e-58;
Matches  383;  Conservative   0; Mismatches   9;  Indels   0;  Gaps   0;

Qy    68 AGACTACTGGGTATCGAACATGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGG 127
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211 AGTCAGTTGGGTATCAGAATGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATCGG 270

Qy   128 ACAAAAGAGGATGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAA 187
            ||||||||||||||||| ||||||||||||||||||||||||| ||||||||||||||
Db   271 ACAAAAGAGGATGTAAATCAGTGGTTAGAAAGTCATAAGATCGACCAAAAACACAGGGAA 330

Qy   188 ATTTTGACTGAACAACAGACGTGAATGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTT 247
            |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   331 ATTTTGACTGAACAAGACGTGAATGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTT 390

Qy   248 GTTGATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTG 307
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   391 GTTGATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTG 450

Qy   308 CGGAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAAT 367
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   451 CGGAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAAT 510

Qy   368 GCTCCTAAAGACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAGCAAAAACAA 427
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   511 GCTCCTAAAGACCAAACGGTGTCTCAAAAGGAACGTAGAGAAACTTCAAGCAAAAACAA 570

Qy   428 AAGGGTAAAGAGAACCCAGATATGGCTAATCC 459
            |||||||||||||||||||||||||||||||
Db   571 AAGGGTAAAGAGAACCCAGATATGGCTAATCC 602

RESULT 13
AW532890/c
LOCUS       AW532890        541 bp    mRNA        linear   EST 06-MAR-2000
DEFINITION  UI-R-BU0-ang-d-03-0-UI.s1 UI-R-BU0 Rattus norvegicus cDNA clone
            UI-R-BU0-ang-d-03-0-UI 3', mRNA sequence.
ACCESSION   AW532890
VERSION     AW532890.1  GI:7175304
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   MEDLINE   97044477
   PUBMED    8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
```

```
                    strand cDNA and therefore this may represent a bonafide poly A
                    tail. The sequence tag present in the cDNA between the NotI site
                    and the oligo-dT track served to identify it as a clone from the
                    normalized Eye library cDNA Library Preparation: M.B. Soares Lab
                    Clone distribution: clones will be available through Research
                    Genetics (www.resgen.com)
                    Seq primer: M13 Forward
                    POLYA=Yes.

FEATURES            Location/Qualifiers
     source         1..541
                    /organism="Rattus norvegicus"
                    /mol_type="mRNA"
                    /strain="Sprague-Dawley"
                    /db_xref="taxon:10116"
                    /clone="UI-R-BU0-ang-d-03-0-UI"
                    /dev_stage="adult"
                    /lab_host="DH10B (Life Technologies)"
                    /clone_lib="UI-R-BU0"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site 1: Not I; Site_2: Eco RI; The UI-R-BU0
                    library is a subtracted library derived from a mixture of
                    eye and ganglia tissues. For a detailed description of the
                    library from which this clone was derived, please visit
                    our web site at ratest.eng.uiowa.edu. The subtraction has
                    been previously described in (Bonaldo, Lennon and Soares,
                    Genome Research 6:791-806, 1996)
                    TAG_TISSUE=Eye
                    TAG_LIB=UI-R-BU0
                    TAG_SEQ=CATTG"

ORIGIN

Query Match           23.8%;    Score 368.2;  DB 10;  Length 541;
Best Local Similarity 80.0%;    Pred. No. 4.8e-56;
Matches  433;  Conservative   0; Mismatches 108;  Indels   0;  Gaps   0;

Qy   681 CCTTCACAAATACAGCAACAGCCACCAGAGAGAGGATGTCAGATGAAAATTAGCAATGAGG 740
            |||| ||| ||||||||||||||||||||| |||||||||||| ||||| |||||||||
Db   541 CCTTAACAACACAGAGACACCCACAGAGGAGGACATAAAGATGAAATTTAGCAACGAAG 482

Qy   741 TTTTCCGATTGCTTCAGCTTGATGAATTCACGTACCAATGCCACTATTCATTTTGGAG 800
            ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   481 TCTTCCGATTGCTTCAGCTTGATGAACTCACGTACTAATGCCACCATTCATTTTTGGAG 422

Qy   801 TCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATGACCAAGGAAG 860
            |||||||||||||||||||||||||| ||||||||||||||||||||||| |||| ||||
Db   421 TCCAAGACAAACCCAAGGAAGGATTGTTGGCATGGAGTCACCACCCGTCACTAAAGATG 362

Qy   861 CCCTATTACCATTCAATCTGATGAATAAACAAGTATTTGAAGACCATCAAGTCCAAC 920
            ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   361 CCCTCATTGACCACTTCAATCTGATGATAGTCCAGTATTTGAGGTCCATCAAGTCCAAC 302

Qy   921 AAGCAAAGAGTGCATTGCGAGAGCCAAGATTTGTGGAAGTTTTACTGCCAAATAGTACTC 980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AAGCAAAGAGTGCATTGCGAGAGCCAAGATTTGTGGAAGTTTTACTGCCAAATAGTACTC 242

Qy   981 TATCTGACAGATTGTTATTGAAGTGGACATTATTTCCACAATATTCAGAATGCCAATATG 1040
            ||||||||||||||||||||||||||||||||||||||| |||||||||||| |||||||
Db   241 CATCTAATAAATTGTTATTGAGGTGGATATCATTCCACAATATTCAGAATGCCAAAGAAG 182

Qy  1041 ATTACTTCCGATTAAAATGCAAAATTCAACAACAACAAAATATGGGAACAAAGTAAAAAT 1100
            ||| ||||||||||||| ||||||||||||||||||||||||||| ||||||||||||||
Db   181 ATTTCTTCCAATAAAAATGGATTGGCATAAGGATGAAACATGGCAAAATTCCAAAAT 122

Qy  1101 TCTCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAAAAATAAAGTTGATT 1160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACTCGGTCTTTGTGAGATGGTCCGAAAAGTAAGGACATCATAGGGATTAAAGCTGATT 62

Qy  1161 TCAGAGCATTTAAAGCAGATTTAAAACACTGGCAGTCCAGAAAAGCAGAAGAAA 1220
            |||| |||||||| ||||||| ||||||||||||||| ||||| |||||||||||||
Db    61 TCAAAGCTTTTAAGTTAGACCTAAAAAGCACTGGCAGACTTCTAGGAAAAAAAAAAAA 2

Qy  1221 A 1221
           |
```

```
Qy   1070  AACAACAAAATATGGGAACAAGTAAGTAAAAAAATTCTCACTATTTGTGCGAGATGGGACCAGC  1129
                  |||||| |||||||| ||||||                    ||||| ||||||||||| ||||
Db    355  AAAGATAAAATATGGAAACAAGAACAAAACAAAATCTTTCACTGTTTGTAAGAGAAGGGGCTAGC  414

Qy   1130  TCTAGGGACATTACGAAAATA--------------AAGTTGATTTCAGAGACATTTAAAGCA  1177
                  ||||||||||| ||||||||              ||||| |||||||||| ||||||||||
Db    415  TCTAGGGGATATCCTGCCAATTCCAAGCACGGATGTAGATTTCAAGGCATTTTTACAA  474

Qy   1178  GATTTTAAAACACTGCAGAGTCCAGAAAGCAGCAGAGAGAAAATTCAGCAGCAAAAACA  1237
                  |||||| |||| ||||||  ||||||||||||||||| |||||||||| |||||||||||
Db    475  AATTTAAAGTCACTGGTAGCATCTAGAAAAGAGGCTGAAGAAGAGTATGGAATGAAGGCA  534

Qy   1238  AATAAAAAGAAAGAGAGAGGGACCAAAGTTGGTTAAATTATTGACAGGAAATCAAGATTTG  1297
                  |||||| |||||||| ||||||||||| ||| ||||||||  ||||||||||| ||| ||
Db    535  ATGAAGAGGAGAGTGAAGGACTAAAGCTGGTTAAACTTCTCATAGGAARACCGAGACTCA  594

Qy   1298  TTAGGTAATTCATACATGCATGAACAGTACATTCTTGTGTAACAAATAAATGCCACCCGAGATCAA  1357
                  ||||| ||||||||||||||||||||||| |||| ||||| |||||||||||| |||||||||||
Db    595  CTGGATAATTCATACTATGACTGGGTACATTCTTGTAACAAATAAATGCCATCCANAACCA  654

Qy   1358  ACAAAACACT  1367
                  || |||||||
Db    655  AATAAGCACT  664
```

```
RESULT 15
BX473839
LOCUS        BX473839                    572 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION   DKFZp686H08165_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
             DKFZp686H08165 5', mRNA sequence.
ACCESSION    BX473839
VERSION      BX473839.1  GI:3168083
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 572)
  AUTHORS    Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
             Fobo,G., Han,M. and Wiemann,S.
  TITLE      EST (Bahr,A., Lauber,J., Mewes,H.W., Weil,B., et al.)
  JOURNAL    Unpublished (2003)
COMMENT      Contact: MIPS
             MIPS
             Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
             This is the 5' sequence of the clone insert
             Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
             Research Center (DKF2), Email s.wiemann@dkfz- heidelberg.de;
             sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
             consortium of the German Genome Project.
             No s1 sequence available.
             This clone (DKFZp686H08165) is available at the RZPD in Berlin.
             Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
             Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..572
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686H08165"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

Query Match            22.7%;  Score 351.8;  DB 13;  Length 572;
Best Local Similarity  98.1%;  Pred. No. 4.1e-53;
Matches  356;  Conservative  0;  Mismatches  7;  Indels  0;  Gaps  0;

Qy     68  AGACTACTGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGG  127
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         1 A 1

RESULT 14
AL700018
LOCUS        AL700018                    664 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION   DKFZp686O10116_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
             DKFZp686O10116 5', mRNA sequence.
ACCESSION    AL700018
VERSION      AL700018.1  GI:19620551
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 664)
  AUTHORS    Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and
             Wiemann,S.
  TITLE      EST (Duesterhoeft, et al.)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: MIPS
             MIPS
             Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
             This is the 5' sequence of the clone insert
             Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
             Research Center (DKF2); Email s.wiemann@dkfz- heidelberg.de;
             sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
             consortium of the German Genome Project.
             No s1 sequence available.
             This clone (DKFZp686O10116) is available at the RZPD in Berlin.
             Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
             Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..664
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686O10116"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

Query Match            23.2%;  Score 358.6;  DB 9;  Length 664;
Best Local Similarity  73.4%;  Pred. No. 2.4e-54;
Matches  492;  Conservative  0;  Mismatches  160;  Indels  18;  Gaps  2;

Qy    710  GAGGATGTCAAGATGAAATTAGCAATGAGGTTTCCGATTTGCTTCAGCTTGTATGAAT  769
                  |||||| || ||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GTGGACATTAAGATGAAATTCAGCAATGAAGTCTTCCGATTTGCATCAGCTTGTATGAAT  60

Qy    770  TCACGTACCACAATGGCACTATTCATTTTGGAGTCAAAGACAAACCCATGGAAAATTGTT  829
                  |||||||||||||||||||| ||| ||||||| |||||| ||||| ||||||||||||||
Db     61  TCACGCACCAATGGCACCACCATTTTGGAGTCAAGGAGACAAACACCATGGAGAAATTGTT  120

Qy    830  GGGATCAAAGTCACCAATGATCCAAGGAAGCCCTCATTAACCATTCAATCTGATGATA  889
                  |||||||||||||||||||||||||||||      ||| |||| ||| |||||||||||
Db    121  GGTGTGAAAATCACCAGT-----AAGGCTGCCTCATTGACCACTTCAATGTAATGATC  174

Qy    890  AACAAGTATTTGAAGACCATCAAGTCCAACAAGCAAAGAAGTGCATTCGAGAGCCAAGA  949
                  ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    175  AAAAAGTATTTTGAAGAAAGTGAGATCAATGAGGCCAAGAAGTGTATTCGGGGAGCCAAGG  234

Qy    950  TTTGTGGAAGTTTACTGCCAAATAGTACTCTATCTGACAGATTTGTTATTGAAGTGGGAC  1009
                  ||||||||||||| |||||||||||||||| |||||||||| |||||||||| ||||| |
Db    235  TTTGTGGAAGTCCTTCTTGCAGAACAATACACCATCTGACAGATTTGCATTGAAGTTGAT  294

Qy   1010  ATTATTCCACAGTTCTCTGAATGCCAATATGATTACTTCCAGATTAAAAATGCAAAATTAC  1069
                  ||||| | |||||||||||||||| ||||| ||||||||||||||||||||||| |||||
Db    295  ACTATTCCAAAACACTCTATATGTAATGATAAGTAGTATTTCTACATTCAGATGCAAAATTTGT  354
```

```
Db  210 AGTCAGTTTGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGAGATGATTGG 269

Qy  128 ACAAAAGAGGATGTAAATCAGTGGTTAGAAAGTCATAAAGTTGACCAAAAACACAGGGAA 187
        ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db  270 ACAAAAGAGGATGTAAATCAGTGGTTAGAAAGTCATAAAGTTGACCAAGAACACAGGGAA 329

Qy  188 ATTTTGACTGAACAAGACGTGAATGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTT 247
        |||||||||||||||||||||||||| |||||||| |||||||||||||||||||||||
Db  330 ATTTTGACTGAACAAGACGTGAATGGAGCAGTCTTCGAAGTGGTTAAAAAAGAACATCTT 389

Qy  248 GTTGATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTG 307
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  390 GTTGATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTG 449

Qy  308 CGGAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAAT 367
        |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db  450 CGGAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAGGCCCAGTAAAAAT 509

Qy  368 GCTCCTAAAGACCAAACTGTGTCTCAAAAGAACGTAGAGAAACTTCAAAGCAAAAACAA 427
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  510 GCTCCTAAAGACCAAACTGTGTCTCAAAAGAACGTAGAGAAACTTCAAAGCAAAAACAA 569

Qy  428 AAG 430
        |||
Db  570 AAG 572
```

Search completed: April 26, 2004, 01:59:33
Job time : 2817.53 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:     April 25, 2004, 07:15:18 ; Search time 76.8439 Seconds
                                    (without alignments)
                          11186.557 Million cell updates/sec

Title:         US-10-051-835-17
Perfect score: 1549
Sequence:      1 aaccatcagatatttcag........ctctaaatcttaccatcaa 1549

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      682709 seqs, 277747446 residues

Total number of hits satisfying chosen parameters:      1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued_Patents_NA:*
            1: /cgn2_6/ptcdata/2/ina/5A_COMB.seq:*
            2: /cgn2_6/ptcdata/2/ina/5B_COMB.seq:*
            3: /cgn2_6/ptcdata/2/ina/6A_COMB.seq:*
            4: /cgn2_6/ptcdata/2/ina/6B_COMB.seq:*
            5: /cgn2_6/ptcdata/2/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/ptcdata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Match Length	DB	ID	Description
1		1462.4	94.4	4771	4	US-09-620-312D-121	Sequence 121, App
2		68.6	4.4	7218	1	US-08-232-463-14	Sequence 14, Appl
3		52.4	3.4	19124	2	US-08-487-826B-13	Sequence 13, Appl
4		47.4	3.1	6326	4	US-10-204-708-58	Sequence 58, Appl
5	c	47		1664976	4	US-08-916-421B-1	Sequence 1, Appli
6	c	46.2	3.0	5562	4	US-10-204-708-63	Sequence 63, Appl
7	c	45.6	2.9	832	4	US-09-621-976-2813	Sequence 2813, Ap
8	c	45.4	2.9	396	4	US-09-640-173-53	Sequence 53, Appl
9	c	45.4	2.9	396	4	US-09-713-550-53	Sequence 53, Appl
10		45.4	2.9	1248	2	US-08-897-340-5	Sequence 5, Appli
11		45.4	2.9	1248	3	US-09-252-329-5	Sequence 5, Appli
12		45	2.9	6243	4	US-09-056-075-1	Sequence 1, Appli
13	c	44.8	2.9	832	4	US-09-621-976-2813	Sequence 2813, Ap
14	c	44.4	2.9	10467	4	US-10-204-708-2	Sequence 2, Appli
15	c	44.2	2.9	6656	4	US-10-204-708-75	Sequence 75, Appl
16	c	43.6	2.8	2447	2	US-09-014-969-14	Sequence 14, Appl
17		43.6	2.8	5152	4	US-10-204-708-47	Sequence 47, Appl
18	c	42.2	2.7	1606	4	US-09-601-198-27	Sequence 27, Appl
19	c	42.2	2.7	1929	4	US-09-543-681A-3758	Sequence 3758, Ap
20	c	42.2	2.7	786431	4	US-09-751-389-3	Sequence 3, Appli
21	c	41.8	2.7	6866	4	US-10-204-708-20	Sequence 20, Appl
22	c	41.6	2.7	2223	1	US-08-257-073-4	Sequence 4, Appli
23	c	41.2	2.7	11049	4	US-10-204-708-22	Sequence 22, Appl
24		41.2	2.7	580073	4	US-08-545-528D-1	Sequence 1, Appli
25	c	41.2	2.6	658	3	US-08-998-416-595	Sequence 595, App
26	c	40.8	2.6	240	1	US-08-628-417-6	Sequence 6, Appli
27		40.8	2.6	1405	1	US-07-710-361-3	Sequence 3, Appli
28	c	40.8	2.6	2210	1	US-07-710-361-2	Sequence 2, Appli
29		40.6	2.6	5562	4	US-10-204-708-64	Sequence 64, Appl
30		40.6	2.6	5433	3	US-08-929-329-1	Sequence 1, Appl
31		40.6	2.6	640681	4	US-09-790-988-1	Sequence 1, Appli
32		40.4	2.6	1696	4	US-09-835-811-1	Sequence 1, Appli
33	c	40.4	2.6	6669	4	US-10-204-708-6	Sequence 6, Appli
34	c	40.4	2.6	11015	4	US-10-204-708-55	Sequence 55, Appl
35	c	40.2	2.6	5852	1	US-07-867-106-2	Sequence 2, Appl
36		40	2.6	4949	3	US-09-138-024-22	Sequence 22, Appl
37		40	2.6	4949	4	US-09-404-066-22	Sequence 22, Appl
38		40	2.6	4949	4	US-09-573-322-22	Sequence 22, Appl
39	c	40	2.6	6070	1	US-10-204-708-10	Sequence 10, Appl
40		39.8	2.6	2394	4	US-09-800-729-33	Sequence 33, Appl
41	c	39.6	2.6	10619	4	US-10-204-708-4	Sequence 4, Appl
42	c	39.4	2.5	8607	4	US-10-204-708-72	Sequence 72, Appl
43	c	39.4	2.5	9347	4	US-10-204-708-36	Sequence 36, Appl
44		39.2	2.5	2671	6	5168051-9	Patent No. 5168051
45		39.2	2.5	5181	1	US-08-257-073-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-620-312D-121

```
; Sequence 121, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
;  APPLICANT: Liu, Chenghua
;  APPLICANT: Tang, Y. Tom
;  APPLICANT: Asundi, Vinod
;  APPLICANT: Zhang, Jie
;  APPLICANT: Ren, Feiyan
;  APPLICANT: Chen, Rui-hong
;  APPLICANT: Zhao, Qing A.
;  APPLICANT: Wehrman, Tom
;  APPLICANT: Xue, Aidong J.
;  APPLICANT: Yang, Yonghong
;  APPLICANT: Wang, Jian-Rui
;  APPLICANT: Zhou, Ping
;  APPLICANT: Ma, Yunqing
;  APPLICANT: Wang, Dunrui
;  APPLICANT: Wang, Zhiwei
;  APPLICANT: John Tillinghast
;  APPLICANT: Drmanac, Radoje T.
;  TITLE OF INVENTION: No. 65696621 Nucleic Acids and
;  TITLE OF INVENTION: Polypeptides
;  FILE REFERENCE: 784CIP2B
;  CURRENT APPLICATION NUMBER: US/09/620,312D
;  CURRENT FILING DATE:  2000-07-19
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 1105
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 121
;  LENGTH: 4771
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (3023)..(4771)
US-09-620-312D-121

Query Match             94.4%;  Score 1462.4;  DB 4;  Length 4771;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 1463;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    86 ATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGACAAAGAGAGGATGTAAT 145
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGACAAAGAGAGGATGTAAAT 60
```

```
Qy   146  CAGTGGTTAGAAGTCATAGATGACCAAAATGACCAAAAACACAGGGAAATTTTGACTGAACAGAGAC  205
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  CAGTGGTTAGAAGTCATAGATGACCAAAATGACCAAAAACACAGGGAAATTTTGACTGAACAAGAC  120

Qy   206  GTGAATGGAGCAGTCTTGAAGTCGTTAAAAAAAGAACATCTTGTTGATATGGCATCACA  265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  GTGAATGGAGCAGTCTTGAAGTCGTTAAAAAAAGAACATCTTGTTGATATGGGCATCACA  180

Qy   266  CATGGACCAGCTATTCAAATAGAGAACTATTCAAAGAATTGCGGAAAACAGCCATTGAA  325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  CATGGACCAGCTATTCAAATAGAGAACTATTCAAAGAATTGCGGAAAACAGCCATTGAA  240

Qy   326  GATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAGACCAAACT  385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAGACCAAACT  300

Qy   386  GTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACAAAAGGGTAAAGAGAACCCA  445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  GTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACAAAAGGGTAAAGAGAACCCA  360

Qy   446  GATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAGGTTCTAAGTCACTAAAAGTT  505
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  GATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAGGTTCTAAGTCACTAAAAGTT  420

Qy   506  GAGCTCATAGAAGATAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA  565
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  GAGCTCATAGAAGATAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA  480

Qy   566  TGTGTATCATATCCATTTGATGAATTCAGTAGTCCATATCGTTACAAGTTGGATTTTAGT  625
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  TGTGTATCATATCCATTTGATGAATTCAGTAGTCCATATCGTTACAAGTTGGATTTTAGT  540

Qy   626  CTACAGCCTGAAACAGGACCAGCCAATCTCATTGATCCGATACATGAATTCAAAGCCTTC  685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  CTACAGCCTGAAACAGGACCAGCCAATCTCATTGATCCGATACATGAATTCAAAGCCTTC  600

Qy   686  ACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGCAATGAGGTTTTC  745
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  ACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGCAATGAGGTTTTC  660

Qy   746  CGATTTGCTTCAGCTTCAGTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAA  805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  CGATTTGCTTCAGCTTCAGTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAA  720

Qy   806  GACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAAAGTCACCAAGGAAGCCCTC  865
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  GACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAAAGTCACCAAGGAAGCCCCTC  780

Qy   866  ATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCAACAAGCA  925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  ATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCAACAAGCA  840

Qy   926  AAGAAGTGCATTCGAGAGCCAAGATTGTGGAAGTTTACTGCAAATAGTACTCCTATCT  985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  AAGAAGTGCATTCGAGAGCCAAGATTGTGGAAGTTTACTGCAAATAGTACTCCTATCT  900

Qy   986  GACAGATTGTTATTGAAGTGGACATTTCCAAGTTCTCTGAATGCCAATATGATTAC  1045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  GACAGATTGTTATTGAAGTGGACATTTCCAAGTTCTCTGAATGCCAATATGATTAT  960

Qy  1046  TTCCAGATTAAAATGCAAAATTACAACAACAAAATATGGAACAAGTAAAAAATTCTCA  1105
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  TTCCAGATTAAAATGCAAAATTACAACAACAAAATATGGAACAAGTAAAAAATTCTCA  1020

Qy  1106  CTATTTGTGCGAGATGGGACCTCAAGGACATTACGAAAAAATAAAGTTGATTTCAGA  1165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  CTATTTGTGCGAGATGGGACCTCAAGGACATTACGAAAAATAAAGTTGATTTCAGA  1080

Qy  1166  GCATTTAAAGCAGATTTTAAAACACTGGCAGTCCAGAAAGCAGCAGAGAAAATTC  1225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  GCATTTAAAGCAGATTTTAAAACACTGGCAGTCCAGAAAGCAGCAGAGAAAATTC  1140

Qy  1226  AGAGCAAAAACAAATAAAAAGAAAAGAAGAGAGGACCAAAGTTGGTTAAATTATTGACAGGA  1285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  AGAGCAAAAACAAATAAAAAGAAAAGAAGAGAGGACCAAAGTTGGTTAAATTATTGACAGGA  1200

Qy  1286  AATCAAGATTTGTTAGATAATTCATCTATGAACAGTACATTCTTGTAACAAATAAATGC  1345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  AATCAAGATTTGTTAGATAATTCATCTATGAACAGTACATTCTTGTAACAAATAAATGC  1260

Qy  1346  CACCCAGATCAAACAAACAACACTTAGATTTCCTGAAGGAAATTAAATGGTTGCTGTATTG  1405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  CACCCAGATCAAACAAACAACACTTAGATTTCCTGAAGGAAATTAAATGGTTGCTGTATTG  1320

Qy  1406  GAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAAAGACCCGAGTA  1465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  GAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAAAGAAGCCCGAGTA  1380

Qy  1466  GCAAACCTTCACTTTCCAAGTGTATATGTTAGAACACAGAAAACCACCACCAAATGAGACGATT  1525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  GCAAACCTTCACTTTCCAAGTGTATATGTTAGAACACAGAAAACCACCACCAAATGAGACGATT  1440

Qy  1526  TCTACTCTAAATCTTTACCATCAA  1549
          ||||||||||||||||||||||||
Db  1441  TCTACTCTAAATCTTTACCATCAA  1464
```

RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
GENERAL INFORMATION:
 APPLICANT: DORNER, F.
 APPLICANT: SCHEIFLINGER, F.
 APPLICANT: FALKNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
　ADDRESSEE: Foley & Lardner
　STREET: 1800 Diagonal Road, Suite 500
　CITY: Alexandria
　STATE: VA
　COUNTRY: USA
　ZIP: 22313-0299
 COMPUTER READABLE FORM:
　MEDIUM TYPE: Floppy disk
　COMPUTER: IBM PC compatible
　OPERATING SYSTEM: PC-DOS/MS-DOS
　SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
　APPLICATION NUMBER: US/08/232,463
　FILING DATE:
　CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
　APPLICATION NUMBER: US/07/935,313
　FILING DATE:
 APPLICATION NUMBER: EP 91 114 300.6
　FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
　NAME: BENT, Stephen A.
　REGISTRATION NUMBER: 29,768
　REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 TELECOMMUNICATION INFORMATION:
　TELEPHONE: (703)836-9300
　TELEFAX: (703)683-4109
　TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
　LENGTH: 7218 base pairs
　TYPE: nucleic acid
　STRANDEDNESS: single
　TOPOLOGY: linear
 IMMEDIATE SOURCE:
　CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match   4.4%;   Score 68.6; DB 1; Length 7218;
Best Local Similarity 3.6%;   Pred. No. 2.7e-07; Indels 143; Gaps 0;
Matches 14; Conservative 234; Mismatches 234;

Qy   46 ATATATGTAAAAGAAATTTGCAAGACTACTGGGTATCAGAATGGCAAAGCAACTTAACCT 105
              || || |||||| || : : :: : ::: : ::
Db 1457 AAAGAGATAGAGAGAATTTGGTACRCRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1398

Qy   106 TCCAGAAAATACAGATGATTGGACAAAAGAGGATGTAAATCAGTCGGTTAGAAAGTCATAA 165
              :::::: ||: ::: ::::::::::: :
Db 1397 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1338

Qy   166 GATTGACCAAAAACACAGGGAAATTTGACTGAACAAGACGTGAACAGGCAGTCTTGAA 225
              :: ::::: ::::::::   : : ::: : : ::: :: : :::
Db 1337 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1278

Qy   226 GTGGTTAAAAAAAAGAACATCTTGTTGATATGGGCATCACACATGGACCAGCTATTCAAAT 285
              :::::::::::: :   : : ::: :::
Db 1277 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1218

Qy   286 AGAAGAACTATTCAAAGAAATTGCGGAAAACAGCCATTGAAGATTCGATTCAGACATCTAA 345
              ::::::::: : :::: :::::: :: :
Db 1217 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1158

Qy   346 GATGGGAAAGCCCAGTAAAAATGCTCCTAAAGACCAAACTGTGTCTCAAAAGGAACGTAG 405
              :: ::::::: : ::::   :: ::: : :::::
Db 1157 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1098

Qy   406 AGAAACTTCAAAGCAAACAAAAACAAAAGGGTAAA 436
              :::::::::::::::::::
Db 1097 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1067

RESULT 3
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellems, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Knobbe Martens Olson & Bear
;   STREET: 620 Newport Center Drive 16th Floor
;   CITY: Newport Beach
;   STATE: California
;   COUNTRY: US
;   ZIP: 92660
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/487,826B
;   FILING DATE: 10-SEP-1993
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Israelsen, Ned
;   REGISTRATION NUMBER: 29,655
;   REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (619) 235-8550
;   TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 19124 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match   3.4%;   Score 52.4; DB 2; Length 19124;
Best Local Similarity 48.3%;   Pred. No. 0.0036;
Matches 146; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

Qy 1062 AAAATTACAACACACAAAATATGGGAACAAAGTAAAAAATTCTCACTATTTGTCGCGAGATG 1121
              |||| ||| || |||||||||||
Db 15597 AAAAAATTAAAAAAATTTATATAAAATGATTATAAAAAAAATAAAAACAAA 15656

Qy 1122 GGACCAGCTCTAAGGACATTACGAAAAAATAAAGTGATTCAGAGCAATTAAAGCAGATT 1181
              ||| |||| |||||||||
Db 15657 AGAGAAAAAAAAACATTAAAAAAAATATATCATAAAAAACAAAAAAAAAG 15716

Qy 1182 TTAAAACACTGGCAGAGTCCAGAAAAGCAGCGAGAGAAAAATTCAGAGCAAAAAACAAATA 1241
              ||||
Db 15717 AAAAAAATATTATTAAAATAAAATATATCATAAAAAAAAATTAAAAAATGTT 15776

Qy 1242 AAAAAGAAAGAGAGGGACCAAAGTTGGTTAAATTATTGACAGGAAATCAAGATTTGTTAG 1301
              || |||
Db 15777 AAAAAAAAAATATATCATAAAATAAAATAAAAAATTTTATTAACAAAATAAAATAAATAATAAAT 15836

Qy 1302 ATAATTCATATCATGACAGTACATTCTTGTAACAAATAAATGCACCCAGATCAAACAA 1361
              || |||
Db 15837 AAAAAAATTTAATTAAATAAAAAAAATATAAAAAAAATTTAATTAAATAAAAAAAA 15896

Qy 1362 AA 1363
              ||
Db 15897 AA 15898

RESULT 4
US-10-204-708-58/c
; Sequence 58, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 1001905B.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 58
; LENGTH: 6326
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-58

Query Match   3.1%;   Score 47.4; DB 4; Length 6326;
Best Local Similarity 44.9%;   Pred. No. 0.04;
Matches 260; Conservative 0; Mismatches 316; Indels 3; Gaps 2;

```
Qy    803 AAGACAAACCCATGGAAAAATTGTTGGCATCAAAGTCACCAATGATACCAA-GGAAGC 861
              ||||||||||  ||| ||||||||||  ||||  | |||| |||||||||  ||||||
Db   2436 AAACAAAAATCTAACTTTATTTTTTTCACACAACAAATTTCCCAATTATCTCAATACGAAT 2377

Qy    862 CCTCATTAACCATTCAATCTGATGATAAACAAGTATTTGAAGACCATCAAGTCCAACA 921
              ||  |||||| ||| |||  | |||||| ||||  | |||  |||  || |||| ||
Db   2376 CTTTTTTTCTTTTTTAATACGATACTATATTTATAAACATCAATACTACATTTATATA 2317

Qy    922 AGCAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTTACTGCCAAATAGTACTCT 981
              |||  |   ||||   ||||||||||  |||||  | | |||||  |||| |||| |
Db   2316 ACTAAAAAAACAAAAAAAATTCATATAATTTTCTTTTACTTAACTAACAAATATTAAATA 2257

Qy    982 ATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAATATGA 1041
              |||| ||  || || || ||||| | ||  |||||  |||| |||||  |||  || |
Db   2256 TCATATATCAAACACAAAACATTACAAAAACCAACAAATATGGGAACAAAGTAAAAAATT 2197

Qy   1042 TTACTTCCAGATTAAAATGCAAAATTACAACAACAAAAATATGGGAACAAAGTAAAAAATT 1101
              ||  || ||  || |||  ||||| |||  |||||  | ||||  || |||| ||||
Db   2196 TCGTATCCTATTTTAAAAATAAATCGCTAATAATAACATAATAATATAAAATAAAATTAAA 2137

Qy   1102 CTCACTATTTGTGCGGAGATGGGACCAGTCTAAGGACATTACGAAAAAATAAAGTTGATTT 1161
              ||||| |||| | |||||| |||||||| ||||| || |||| |||||  |||| |||||
Db   2136 ATAAAAACCTAACCACCCAAAACTATAAACAAAAAACTAACTTACCCTACCAATAATAT 2077

Qy   1162 CAGAGCATTTAAAGCAGATTTAAAACTGGCAGAGTCCAGAAAAGCAGCAGAAGAAAA 1221
              ||| || |||  | ||||||||||| ||  |  |||||  ||| ||||| ||  |||
Db   2076 TAAAAAACTTAAACTAAACTATAAACCAT--CCAACTACATCCTCCAAACCTAACGAAA 2019

Qy   1222 ATTCAGAGCAAAACAAATAAAAAGAAGAGAGAGACCAAAGTTGGTTAAATTATTGAC 1281
              ||| || ||||||||| |  |||| | ||||| |||||  |||| |||| ||  |||
Db   2018 CTTTCTCTCAAACAAAACAAAAACGCAAACACAAACTTCAAATACGTATTTACTTCGATAA 1959

Qy   1282 AGGAAATCAAGATTTGTTAGATATTCATACTATGAACAGTACATTCCTGTAACAAATAA 1341
              ||| ||||| |||| |||  |||||  || ||  ||| | |  |||| | |||| ||||
Db   1958 TAATATTTAATTTAATTTAACCAAAACATTAACCTACAAAACAAAAAATAAAARATAA 1899

Qy   1342 ATGCCACCCAGATCAAACAAAACAACTTAGATTTCCTGAA 1380
              ||| |||  | || |||  |||| ||  |||||| ||| ||
Db   1898 ATAAAAATAAATATAACAAAATCCTAATATCTACTTAA 1860
```

RESULT 5
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103398)..(103398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g

Query Match   3.0%; Score 47; DB 4; Length 1664976;
Best Local Similarity   48.1%; Pred. No. 0.38;
Matches 164; Conservative 0; Mismatches 175; Indels 2; Gaps 1;

```
Qy    1050  AGATTAAAATGCAAAATGCACACAACAAATATGGGACAACAAGTAAAAAATTCTCACTAT 1109
              ||||||||| | ||||| |||||   ||||| |||| |||| ||| ||||| | ||||
Db  844227  AAATTAAAAGACTAATTCAAAAATTAGAGAATTAAAAAATATAACAA--CGCTTTTA 844170

Qy    1110  TTGTGCGGATGGGACCAGCTCTAAGGACATTACGAAAAAATAAAGTTGATTTCAGAGCAT 1169
              | |||| || ||||  | ||||  | |||||||  | |||  | ||| ||||||||||
Db  844169  TTTTGTTAGATGAGAAGTTTTTAGACATCTTTTTACAAAAAGATTTGGATGAAATAATAA 844110

Qy    1170  TTAAAGCAGATTTTAAAACACTGCAGAGTCCAGAAAAAGCAGCAAGAAAAATTCAGAG 1229
              ||| |  |||| |||   | || | |||| || |  || |||| | |||| |||||
Db  844109  ATGAATATAAGGATTTTGACTTTATATTTTACTACACTGGAGAAGAAGAGAAAAGAAAAAC 844050

Qy    1230  CAAAAACAAATAAAAAAGAGAAGGAGGGACCAAAGTTGGTTAAATTATTGACAGGAAATC 1289
              ||||| ||||||||||||| | || |||| ||||||| | ||||||||  | | |||
Db  844049  CTAAAGAAGTAAAAAAAGAGATTAAAAAGAAACTGAAGAGAAAAATAGAAAAAGAAAAAA 843990

Qy    1290  AAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTAACAAATAAATGCCACC 1349
              ||||| || || || ||||  | | | |||| ||| |||||  ||| ||||  |||
Db  843989  TAGAATTTGTTAAGAAGAAGAAAAGAACAAATTTATAAAAAAATCTGATGAAGATGTTG 843930

Qy   .1350  CAGATCAAACAAAACACTTAGATTTCCTGAAGGAAATAAA 1390
              |||| || |||| |  ||||| ||||  ||| |||||
Db  843929  AAGAGAAATTAAAACAACTAATTTCCAAAGAAGAGAAAAAA 843889
```

RESULT 6
US-10-204-708-63/c
; Sequence 63, Application US/10204708
; Patent No. 677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 1001905B.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 1004382G.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 63
; LENGTH: 5562
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-63

Query Match   3.0%; Score 46.2; DB 4; Length 5562;
Best Local Similarity   49.2%; Pred. No. 0.075;
Matches 121; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```
Qy    1145  AAAAATAAAGTTGATTTCAGAGCATTTCAGAGCATTTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGA 1204
              ||||||| ||||| || | | ||||| | | || | | ||||  ||| | ||||||||
Db    1625  AAAAAAAATATAAAAAAAAATTAATAAAATTAATAAACTTTCTTAAACAAAAAAAAATAAAAAA 1566

Qy    1205  AAAGCAGCAGAAGAAAATTCAGAGCAAAACAATAAAAAGAAGAGGGACCAAAG 1264
              |||| | | | ||   |||||| | |||| | || || | |||||
Db    1565  AAAACACTATAAAAAAACAAAAAACAAAAATATAAAAAAAAAAATCTAACTAACAAAAATAA 1506
```

```
Qy    1265  TTGGTTAAATTATTGACAGGAAATCAAGATTTGTTTAGATAATTCATACTATGAACAGTAC  1324
                  |||||||||||||||||||||  |||||||||||||||   ||||  ||
Db    1505  AAATAATAAAAATAAATAAATAAAAACTAAACTAAAATAAAAAATATAAATAT          1446

Qy    1325  ATTCTTGTAACAAATAAAATGCCACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAA  1384
                  ||||   |||||||  |||||||||||||||||||||||||||
Db    1445  AATCAAAAAAAAAAATCATTCCTAACCGAAAAAAAAATAATAATAATATTAAAATA        1386

Qy    1385  ATTAAA  1390
             ||||||
Db    1385  ATAAAA  1380
```

RESULT 7
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match   2.9%; Score 45.6; DB 4; Length 832;
Best Local Similarity   15.3%; Pred. No. 0.053; Indels   1; Gaps   1;
Matches   44; Conservative 132; Mismatches 111; Indels 111; Gaps   1;

```
Qy     810  AACCCCATGGGAGAAATTGTTGGCATCAAAGTCACCAATGATACCAAGGAAGCCCTCATTA  869
                 : :: : ||:||:|:|:| :|| ||:|: ::: :: : ::|
Db      87  MMWMYWGWRRSYNAMWTRTWTGYAYYRSMMYWRYRCWKKKAYYRKTTCYSSKGWTTWWKR   146

Qy     870  ACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCAACAAGCAAAGA  929
                 :|||  ::  : :::: ||:| : :: |||
Db     147  WKKAWTTWWWKKTYWAAATRYWWMCWTKRWRASWWYCWWWGKARKWSTWRKSRSYASAR   206

Qy     930  AGTGCATTCGAGAGCCAAGATTTGTGGAGTTTTACTGCCAAATAGTACTCTATCTGACA  989
                 : |::||:|| :: : :::|:|  | :::|  ::|  ::: :|
Db     207  SAKRCCYSCSWGAMSWKYMWRMWRGWATGAGMKAWRASCMWRRKYAGKSKTSYKSMWM   266

Qy     990  GATTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAATATGATTACTTCC  1049
                 : ::|:|| :: :::|| ::| ::: ||:::| ||:|
Db     267  CWTRSWKYCYTKARWTGYYCYRKGG-MWGKGRWYASKKYMWKRWWWCWARMYRYSTGTR  325

Qy    1050  AGATTAAAAATGCAAAATTACAACAACAAAAATATGGGAACAAAGTAAAA  1097
                 :::|:| |  : :: ::: :|
Db     326  ASMWNRRWYTMMMKWWKYAWARAAWRAAWRWWAMWWAWRRACAAAATATAA  373
```

RESULT 8
US-09-640-173-53/c
; Sequence 53, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173

───

; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-640-173-53

Query Match   2.9%; Score 45.4; DB 4; Length 396;
Best Local Similarity   38.3%; Pred. No. 0.046; Indels   0; Gaps   0;
Matches 142; Conservative   0; Mismatches 229; Indels   0; Gaps   0;

```
Qy     923  GCAAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAGTTTTACTGCCAAATAGTACTCTA   982
                 ||||
Db     372  GCAAGNNNNNNAANNNNAGANNNNNNNNNNNNNNNNCTCGCNNNCCCNNTNNNNNAANN   313

Qy     983  TCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAATATGAT  1042
                 ||| ||||| |||||||||||||||||||||||||||||||||||
Db     312  ANTATNAAGNANNATNNNNNNTNNGNTNNNTNNTNTATNTTTCTNTTTCTTTTTNTGAA   253

Qy    1043  TACTTCCAGATTAAAATGCAAAATTACAACAACAAAATATGGGAACAAAGTAAAAAATTC  1102
                 ||| ||||||| ||||  |||||||||||||||||||||||||
Db     252  TTAAAANAAGNAAAANAAAAAAAANNTAAAAAAAAAAAAAAAAAAAAAAAA          193

Qy    1103  TCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAAAATAAAGTTGATTTC  1162
                 ||| |||  || |||||||||||||||||||||||||
Db     192  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA          133

Qy    1163  AGAGCATTAAAGCAGATTTTAAAACACTGGCGAGCTCCAGAAAAGCAGCAGAAGAAAAA  1222
                 |||||||||||||||||||||||||||||||||||
Db     132  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA           73

Qy    1223  TTCAGAGGCAAAACAAATAAAAAGAGAGGACCAAAGTTGGTTAAATTATTGACA  1282
                 ||| ||||||||||||||||||||||||||||||||||
Db      72  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA            13

Qy    1283  GGAAATCAAGA  1293
                 |||||||||
Db      12  AAAAAAAAAAA   2
```

RESULT 9
US-09-713-550-53/c
; Sequence 53, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-713-550-53

Query Match   2.9%; Score 45.4; DB 4; Length 396;
Best Local Similarity   38.3%; Pred. No. 0.046; Indels   0; Gaps   0;
Matches 142; Conservative   0; Mismatches 229; Indels   0; Gaps   0;

Qy   923 GCAAGAAGTGCATTCGAGAGCCAAGATTTGTGAAGTTTTACTGCCAAATAGTACTCTA 982
               ||||
Db   372 GCAAGNNNNNAANNNAGANNNNNNNNNNNNNNNNNCTCGCNNNCCCNNTNNNNNAANN 313

Qy   983 TCTGACAGATTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAATATGAT 1042
               |||| |||||
Db   312 ANTATNAAGNANNATNNNNNNNTNNGNTNNNTNNTNTATNTTCTNTTCTTTTTTNTGAA 253

Qy   1043 TACTTCCAGATTAAAATGCAAAATTACAACAACAAAATATCGGGACAAAGTAAAAATTC 1102
               ||| |||||
Db   252 TTAAAANAAGGNAAANAAANAANNTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 193

Qy   1103 TCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACCTCTAAGGAAAAATAAAGTTGATTTC 1162
               |||| |||||
Db   192 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 133

Qy   1163 AGAGCATTTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAAAGCAGCAGAAGAAAA 1222
               ||||| |||||
Db   132 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 73

Qy   1223 TTCAGAGCAAAAACAAATAAAAAAGAGAGGGACCCAAAGTTGGTTAAATTATTGACA 1282
               |||||||||| ||||||
Db   72 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 13

Qy   1283 GGAAATCAAGA 1293
               ||| ||
Db   12 AAAAAAAAAA 2

RESULT 10
US-08-897-340-5
; Sequence 5, Application US/08897340
; Patent No. 5955306
; GENERAL INFORMATION:
;   APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.
;   TITLE OF INVENTION: Weight Control Pathway Genes and Uses
;   TITLE OF INVENTION: Therefor
;   NUMBER OF SEQUENCES: 36
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: LAHIVE & COCKFIELD, LLP
;   STREET: 28 State Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: USA
;   ZIP: 02109
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/897,340
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/715,032
;   FILING DATE: 17-SEP-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Silveri, Jean M.
;   REGISTRATION NUMBER: 39,030
;   REFERENCE/DOCKET NUMBER: MNI-005CP
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617)227-7400
;   TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1248 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
US-08-897-340-5

Query Match   2.9%; Score 45.4; DB 2; Length 1248;
Best Local Similarity   51.9%; Pred. No. 0.069;
Matches   97; Conservative   2; Mismatches   88; Indels   0; Gaps   0;

Qy   1059 TGCAAAATACAACAACAAAATATGGGACAAAGTAAAAATTCCACTATTTGTGCGAG 1118
               |||| ||||||
Db   1062 TGAAGAATTTTTTCAAGTAAACATTTGAAGGCTGTAGACATTTTGCATTTTGTACCTG 1121

Qy   1119 ATGGGACCAGCTCTAAGGACATTACGAAAAATAAGTTGATTTCAGAGCATTTAAAGCAG 1178
               ||||| ||||
Db   1122 CAAGTGCCATCTTTAAGGGGGAAMTACATGAAGTCACCGTTACCAGTAACTTGATGTGTA 1181

Qy   1179 ATTTTAAAACACTGGCAGAGTCCAGAAAAGCAGCAGAGAAAATTCAGAGCAAAAACAA 1238
               ||||| |:|||||
Db   1182 TATTAATAAAGTAATTCAGTCMAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1241

Qy   1239 ATAAAAA 1245
               |||||
Db   1242 AAAAAAA 1248

RESULT 11
US-09-252-329-5
; Sequence 5, Application US/09252329
; Patent No. 6147192
; GENERAL INFORMATION:
;   APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.
;   TITLE OF INVENTION: Weight Control Pathway Genes and Uses
;   TITLE OF INVENTION: Therefor
;   NUMBER OF SEQUENCES: 36
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: LAHIVE & COCKFIELD, LLP
;   STREET: 28 State Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: USA
;   ZIP: 02109
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/252,329
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/897,340
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Silveri, Jean M.
;   REGISTRATION NUMBER: 39,030
;   REFERENCE/DOCKET NUMBER: MNI-005CP
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617)227-7400
;   TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1248 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
US-09-252-329-5

Query Match   2.9%; Score 45.4; DB 3; Length 1248;
Best Local Similarity   51.9%; Pred. No. 0.069;
Matches   97; Conservative   2; Mismatches   88; Indels   0; Gaps   0;

Qy   1059 TGCAAAATTACAACAACAAAATATGGGACAAAGTAAAAATTCCTCCACTATTTGTGCGAG 1118
               |||| ||||||
Db   1062 TGAAGAATTTTTTCAAGTAAACATTTGAAGGCTGTAGACATTTTGCATTTTGTACCTG 1121

Qy   1119 ATGGGACCAGCTCTTAAGGACATTACGAAAAATAAAGTTGATTTCAGAGCATTTAAAGCAG 1178
               ||| |||||| |||| || :| || ||||
Db   1122 CAAGTGCCCATCTTTAAGGGGGAAMTACATGAGTCACCGTTACAGTAACTTGATGTGTA 1181

Qy   1179 ATTTTAAAACACTGGCAGTCCGAGAAAAAGCACCAGAAGAAAAATTCAGAGCAAAAACAA 1238
               ||| :| |||| ||||| ::|| ||||||
Db   1182 TATTAATAAAGTAATTCAGTCMAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1241

Qy   1239 ATAAAAA 1245
               | |||||
Db   1242 AAAAAAA 1248

RESULT 12
US-09-056-075-1
; Sequence 1, Application US/09056075
; Patent No. 5955368
; GENERAL INFORMATION:
;   APPLICANT: Johnson, Eric A.
;   APPLICANT: Bradshaw, Marite
;   APPLICANT: Rood, Julian
;   TITLE OF INVENTION: Expression System for Clostridium
;   TITLE OF INVENTION: Species
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Quarles & Brady
;   STREET: 1 South Pinckney Street
;   CITY: Madison
;   STATE: WI
;   COUNTRY: US
;   ZIP: 53701-2113
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/056,075
;   FILING DATE:
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Seay, Nicholas J.
;   REGISTRATION NUMBER: 27386
;   REFERENCE/DOCKET NUMBER: 960296.95238
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 608-251-5000
;   TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 6243 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3770..4013
;   OTHER INFORMATION: /note= "RP4 origin of DNA transfer (oriT) from
;   OTHER INFORMATION: plasmid RP4"
US-09-056-075-1

Query Match   2.9%;   Score 45; DB 2; Length 6243;
Best Local Similarity 46.6%; Pred. No. 0.15;
Matches   144; Conservative   0; Mismatches 165; Indels   0; Gaps   0;

Qy   1029 AATGCCAATATGTATTACTTCCAGATTAAAATGCAAAATTACAACAACAAAATATGGGAC 1088
               |||||||| || ||||||| ||| | || || |||| ||||||| ||||||
Db   1166 AATACAAGCGTAAAAAAATGAGGGTAAAATAAAAATAAAAAAATAAAAAAATAAAAA 1225

Qy   1089 AAAGTAAAAAATTCTCACTATTTGTGCGACGATGGGACCAGCTCTTAAGGACATTACGAAAA 1148
               | |::::::::|| :::: ::|| :::||||||||||||||||||||||||||||
Db   1226 ATAAAAAAAATAAAAAAATAAAAAAAATAAAAAAAAAATAAAAAAAAATAAAAAAAAT 1285

RESULT 13
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;   APPLICANT: Dumas Milne Edwards, J.B.
;   APPLICANT: Jobert, S.
;   APPLICANT: Giordano, J.Y.
;   TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;   FILE REFERENCE: GENSET.054PR2
;   CURRENT APPLICATION NUMBER: US/09/621,976
;   CURRENT FILING DATE: 2000-07-21
;   NUMBER OF SEQ ID NOS: 19335
;   SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

Query Match   2.9%;   Score 44.8; DB 4; Length 832;
Best Local Similarity 12.9%; Pred. No. 0.084;
Matches   51; Conservative 177; Mismatches 164; Indels   4; Gaps   1;

Qy   853 CAAGGAAGCCCTCCATTAACCATTCAATCTGATGATAAACAAGTATTTGAAGACCATCA 912
               |||||| |||||| | :|:||::|:: || :: ::||
Db   394 CAAGAATTGTACCATAATAAATTATATTTGTYWTWNYTWWYTYWTRMWWMKKKAR 335

Qy   913 AGTCCAACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTCTGGAAGTTTACTGCCAAA 972
               ::::|| :::: :::: : | :: :::::
Db   334 RWYYWWKSTYACASRYRKYTWGWWNYNWKRMSTRWYCYMCWKCCMYRGRRCAWYTMARG 275

Qy   973 TAGTACTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATG 1032
               :: :: :: ::|::: :: :::|:: ::||
Db   274 RMNSYAWGKWKSMRSAMSMCTRMYYKKGSTYWTMKCTCATWCYWYWKYWKRMWSKTCWSG 215

Qy   1033 CCAATATGATTACTTCCAGATTAAAATGCAAAATTACAACAACAAAATATGGGAACAAAG 1092
               || :::|:: :: ||:   ::| | :: :|: :|:::::| ::::|:
Db   214 SRGGYMTSYTSTRSYSM---YWASWNYTMCWWNGRWWSTYWYMAWGKGWWRYATTWRRA 159

Qy   1093 TAAAAAATTCTCACTATTTGTCGACGATGGGACCAGCTCTAAGGACATTACGAAAATAA 1152
               ::::|:|:: :: ::| ::|
Db   158 MWWWWAAWTMMWYWWAWCMSSRGGAAMYRRTMMMWGYRYWRKKSYRRTRCAWAYAWKTK 99

Qy   1153 AGTTGATTTCAGAGCATTTAAAGCAGATTTTAAAACACTGGCAGTCCAGAAAAGCAGC 1212
               ::: :: ::: : :| :::::| | :|:::::: :: :|
Db   98 RSTYWCWRWKWKRCMMMMMMAVAYGKTMMMRACWKTRYWRWWAWAMWRMWTMMMMYYWY 39

Qy   1213 AGAAGAAAAATTCAGAGCAAAAACAAATAAAAAAGA 1248
               ::|:::::::: ::::: :::|::::::::::
Db   38 WRAWKRRWWMRKWRSWSMWMWAWGMTRWAARMWWRW 3

Qy   1149 ATAAAGTTGATTTCAGAGCATTTAAAGCAGATTTTAAAAACACTGCAGAGTCCGAAAAAG 1208
               | ||| | || ||||| ||||| :|| |||||| |||||||||| ||||||||
Db   1286 AAAAAAATAAAAAATATAAAAAAAATATAAAAAAAATAAAAAAAATATAAAATATATATAAAAA 1345

Qy   1209 CAGCAGAAGAAAAATTCAGAGCAAAAACAAATAAAAAAGAGAGGGACCAAAGTTGG 1268
               | :| :|:|| | || ||||| |: :| ||| :|| ||| |||||||| |||||
Db   1346 AAATATAAAAAATAAAAAATAAAAAATATAAAAATATAAATATATAAAAT 1405

Qy   1269 TTAAATTATTGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATTC 1328
               |||||:::|:|| :::||||| ::| | | ||:: |:| ||| :||::
Db   1406 ATTTTTTATTTAAAGTTGAAAAAATTTTTTTTATTATTATAATCTTTGAAGAAAAGAA 1465

Qy   1329 TTGTAACAA 1337
               ||| ||||
Db   1466 TATAAAAAA 1474

RESULT 14
US-10-204-708-2/c
; Sequence 2, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 10467
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-2

Query Match   2.9%;   Score 44.4; DB 4;   Length 10467;
Best Local Similarity   48.6%;   Pred. No. 0.26;
Matches 180;   Conservative   0; Mismatches 186; Indels   4; Gaps   2;

Qy   1035 AATATGATTACTTCCAGATTAAAATGCAAAATTACACACAACAAAATATGGGAACAAAGTA 1094
          |||||| |||||||| |||||||| |||| ||| | |||||||||||||| |||||||
Db   8228 AAAAAAATCAATTCAAAAAAAATTAAAAAAATAAAAAAACACGTAAAAAAACAAATTA 8169

Qy   1095 AAAAATTCTCACTATTTGTGCGAGATGGGACCCAGCTCTAAGGACATTACGAAAAATAAAG 1154
          |||| || |||||||||||||||||||||||| || ||||||||| ||||||||||
Db   8168 AAAAAAATAAAACAAAATAAAATAATATATCCAAATATTTCGATAAATCGCAATAAAAATT 8109

Qy   1155 TTGATTTCAGAGGCATTTAAAGCAGATTTAAAAAACACTGGCAGAGTCCAGAAAAGCAGAG 1214
          |||||| |||||| |||||| ||||| |||||||| |
Db   8108 ATAAACTCCAAACAACAAAAAAATTATCAAACCAA---ATATTTTAAATTATTCAAAT 8052

Qy   1215 AAGAAAAATTCAGACGAACAAAAACAAAATAAAAAAAGAAAGAGAGGGACCAAAGTTGGTTAAAT 1274
          || |||||| |||||||||| || |||||||| |||| ||
Db   8051 AAATACTATTTTACAAAAACACATTTAAAACAATAAAAAAAAATTAAAAATATAAA 7992

Qy   1275 TATTGACAGGAAATCAAGATTTGTTAGATAATTC-ATACTATGAACAGTACATTCTTGTA 1333
          |||| || ||| |||||||||||||| |||| ||| | ||||||||||||
Db   7991 TTTTAAAAAATTCAAATATATTTTAATCAAAAAAATATAAAATATAACTAATAAAATACTA 7932

Qy   1334 ACAAATAAATGCCACCCAGATCAAACAAACACTTAGATTTCCTGAAGGAAATTAAATGG 1393
          |||| |||||| |||| |||||||||||| || ||||| |||||| ||||
Db   7931 AACGATAAAAAAACTTTAATATAATATATAAAAAACACTAAATTCTCCAAAAAAAATAACAATT 7872

Qy   1394 TTTGCTGTAT 1403
          ||||
Db   7871 TAAAATATAT 7862

RESULT 15
US-10-204-708-75/c
; Sequence 75, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 75
; LENGTH: 6656
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-75

Query Match   2.9%;   Score 44.2; DB 4;   Length 6656;
Best Local Similarity   50.2%;   Pred. No. 0.25;   Gaps   1;
Matches 135;   Conservative   0; Mismatches 133; Indels   1;

Qy   1009 CATTATTCCACAGTTCTCTGAATGCCAATATGATTACTTCCAGATTAAAATGCAAAATTA 1068
          |||||| ||||| || || ||| |||||||||||
Db   278 CGTCGTTACACTCCAACCTAAACGACAAAATAAAAAAACTATCTCAAATAAATAATAA 219

Qy   1069 CAACACAAAATATGGGACCAAAAGTAAAAAAATTCTCACTATTTGTGCGAGATGGGACCAG 1128
          || |||||||||| |||||| ||| |||||||||||||||||
Db   218 ATAAATAAAAATACGAACAATAATCAAAAAACTTATACTATATTATTCCAAATAAAATAT 159

Qy   1129 CTCTAAGGACATTACGAAAAATAAAGTTGATTTCAGAGCATTTAAAGCAGATTTTAAAAAC 1188
          |||| |||||||| |||||||||| ||||||| ||||||||| |||| |||||
Db   158 CTCTAACCAAATAAAAAAAAAAAATAAATTTTTATATATTTTAAAATAAAC 99

Qy   1189 ACTGGCAGAGTCCAGAAAAGCAGCAGCAGAGAAAATTCAGCAGAAAAATTCAGAGCAGCAAAAACAAAATAAAAAAAGA 1248
          |||||||| |||||| || ||||| ||||||||||||||| ||||
Db   98 -CTATAAATTTATTAATATTAATTACGAAATATAAACCAAACAAAAATATAAAAAAATA 40

Qy   1249 AAGAGAGGGACCAAAGTTGGTTAAATTAT 1277
          ||| ||||||||||| ||||||
Db   39 AAATTTTTAACCTAAATAACTAAAAAAT 11

Search completed: April 26, 2004, 02:11:14
Job time : 81.8439 secs

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    April 24, 2004, 23:47:42 ; Search time 411.576 Seconds
                                  (without alignments)
                           15988.422 Million cell updates/sec

Title:          US-10-051-835-17
Perfect score:  1549
Sequence:       1 aacacatcagatatttcag.........ctctaaatcttaccatcaa 1549

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:    6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      N Geneseq_29jan04:*
                 1: geneseqn1980s:*
                 2: geneseqn1990s:*
                 3: geneseqn2001as:*
                 4: geneseqn2001as:*
                 5: geneseqn2001bs:*
                 6: geneseqn2002s:*
                 7: geneseqn2003as:*
                 8: geneseqn2003bs:*
                 9: geneseqn2003cs:*
                10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                          SUMMARIES
```

Result No.	Score	% Match	Query Length	DB	ID	Description
1	1549	100.0	1549	9	ADE25681	Ade25681 Human cDN
2	1470.4	94.9	4778	4	AAI24129	Aai24129 Probe #14
3	1470.4	94.9	4778	4	ABA69252	Aba69252 Human foe
4	1470.4	94.9	4778	4	AAI49425	Aai49425 Probe #18
5	1470.4	94.9	4778	4	ABA51244	Aba51244 Human bre
6	1470.4	94.9	4778	4	ABA36171	Aba36171 Probe #14
7	1470.4	94.9	4778	4	AAK43354	Aak43354 Human bon
8	1470.4	94.9	4778	4	AAK17545	Aak17545 Human bra
9	1470.4	94.9	4778	4	ABS42976	Abs42976 Human liv
10	1470.4	94.9	4778	5	AAI09703	Aai09703 Probe #96
11	1470.4	94.9	4778	6	ABS17437	Abs17437 Human gen
12	1462.4	94.4	4771	4	AAI58244	Aai58244 Human pol
13	1462.4	94.4	4771	8	ADB48211	Adb48211 Novel hum
14	1462.4	94.4	6233	4	AAS01147	Aas01147 Interfero
15	1043.4	67.4	1045	4	AAK77710	Aak77710 Human imm
16	672	43.4	705	6	ABN60169	Abn60169 Human can
17	636.2	41.1	6289	5	AAL60054	Aal60054 Human Pc0
18	628.4	40.6	5870	4	AAS01146	Aas01146 Interfero
19	564.6	36.4	4185	4	AAI22155	Aai22155 Probe #12
20	564.6	36.4	4185	4	ABA67234	Aba67234 Human foe
21	564.6	36.4	4185	4	AAI47453	Aai47453 Probe #16
22	564.6	36.4	4185	4	ABA34329	Aba34329 Probe #12
23	564.6	36.4	4185	4	AAK41409	Aak41409 Human bon
24	564.6	36.4	4185	4	ABS41002	Abs41002 Human liv
25	564.6	36.4	4185	5	AAI07856	Aai07856 Probe #78
26	564.6	36.4	4185	6	ABS15415	Abs15415 Human gen
27	542.4	35.0	992	7	ABX63726	Abx63726 Human cDN
28	388	25.0	389	8	ACH31668	Ach31668 Human bon
29	361	23.3	644	4	AAS35870	Aas35870 Human car
30	361	23.3	644	4	AAK69748	Aak69748 Human imm
31	361	23.3	644	9	ADE46564	Ade46564 Human car
32	227.6	14.7	604	4	AAS27397	Aas27397 cDNA enco
33	227.6	14.7	604	9	ADB93575	Adb93575 Human cDN
34	138	8.9	356	4	AAS37122	Aas37122 Novel hum
35	73	4.7	251	4	AAS38441	Aas38441 Novel hum
36	71.6	4.6	12718	4	AAS35869	Aas35869 Human car
37	71.6	4.6	12718	4	AAK69747	Aak69747 Human imm
38	71.6	4.6	12718	9	ADE46563	Ade46563 Human car
39	61	3.9	14006	6	ABL33958	Abl33958 Human imm
40	58.6	3.8	6668	6	ABL33697	Abl33697 Human imm
41	58.4	3.8	2000	7	ADA71938	Ada71938 Rice gene
42	58.4	3.8	5750	4	AAS46708	Aas46708 Tumour su
43	58.4	3.8	5750	6	ABL34008	Abl34008 Human imm
44	56.8	3.7	6644	3	AAX33181	Aax33181 Base sequ
45	56.8	3.7	7372	2	AAX33182	Aax33182 Base sequ

```
                         ALIGNMENTS

RESULT 1
ADE25681
ID   ADE25681 standard; cDNA; 1549 BP.
XX
AC   ADE25681;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human cDNA differentially expressed in foam cells #85.
XX
KW   Human; ss; differential expression; foam cell; LPS; lipopolysaccharide;
KW   cardiovascular disease; atherosclerosis.
XX
OS   Homo sapiens.
XX
PN   US2003194721-A1.
XX
PD   16-OCT-2003.
XX
PF   18-SEP-2002; 2002US-00247671.
XX
PR   19-SEP-2001; 2001US-0323784P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Mikita T, Shiffman D, Porter JG, Kaser MR;
XX
DR   WPI; 2003-875398/81.
XX
PT   Combination containing several polynucleotide that are differentially
PT   expressed in foam cells and complements of the polynucleotides, useful
PT   for diagnosing cardiovascular disease or atherosclerosis.
XX
PS   Claim 1; SEQ ID NO 85; 37pp; English.
XX
CC   The invention relates to a combination comprising several polynucleotides
CC   having any one of 127 sequences (S1) such as the sequence of human
CC   calmodulin gene, human mRNA for KIAA0930 protein, leukotriene A4
CC   hydrolase, human CGI-142 protein mRNA, human K+ channel beat 2 subunit
CC   mRNA, etc., and their complements. The cDNAs are differentially expressed
CC   in LPS (lipopolysaccharide)-treated foam cells. Also included are
CC   obtaining an extended or full length gene from a library of nucleic acid
CC   sequences, an expression vector containing the nucleic acids, a host cell
CC   containing the vector, a purified polypeptide appearing as ADE25750 and
CC   ADE25751, producing a protein by culturing the host cell, and a
CC   composition comprising a purified antibody that specifically binds to the
```

```
CC   proteins. The foam cell-expressed nucleic acids are useful for a high
CC   throughput detection of differential expression of one or more
CC   polynucleotides in a sample. The sample is from a subject with
CC   atherosclerosis and comparison with a standard defines early, mid or late
CC   stages of the disorder. The foam cell-expressed nucleic acids are useful
CC   for high throughput screening of a library of molecules or compounds to
CC   identify a ligand which binds a polynucleotide. The library is chosen
CC   from DNA molecules, peptides, proteins and RNA molecules. The protein is
CC   useful for a high throughput screening of library of molecules or
CC   compounds to identify at least one ligand which specifically binds a
CC   protein, for purifying a ligand from a sample for making a antibody. The
CC   foam cell-expressed nucleic acids are useful for diagnosing
CC   cardiovascular disorder. The foam cell-expressed nucleic acids are useful
CC   as elements on a microarray which can be used for detecting related
CC   polynucleotide in a sample, diagnosing cardiovascular disease,
CC   atherosclerosis. The present sequence represents a cDNA whose expression
XX   is upregulated in LPS treated foam cells.
SQ   Sequence 1549 BP; 609 A; 274 C; 287 G; 379 T; 0 U; 0 Other;

Query Match      100.0%;  Score 1549;  DB 9;  Length 1549;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1549;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 AACACATCAGATATTTTCAGCACTAAAAGAGATGGTTTTCCCCACATATATGTAAAAGAAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AACACATCAGATATTTTCAGCACTAAAAGAGATGGTTTTCCCCACATATATGTAAAAGAAA 60

Qy     61 ATTTGCAAGACTACTGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGA 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATTTGCAAGACTACTGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGA 120

Qy    121 TGATTGGACAAAAGAAGAGATGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGATTGGACAAAAGAAGAGATGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACA 180

Qy    181 CAGGGAAATTTTGACTGAACAAGAACGTGAAATGAGCAGTCTTGAAGTGGTTAAAAAAAGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGGGAAATTTTGACTGAACAAGAACGTGAAATGAGCAGTCTTGAAGTGGTTAAAAAAAGA 240

Qy    241 ACATCTTGTTGATATGGGCATCACACATGACCAGCTATTCAAATGAAGAACTATTCAA 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACATCTTGTTGATATGGGCATCACACATGACCAGCTATTCAAATGAAGAACTATTCAA 300

Qy    301 AGAATTGCGGAAAACAGCCATTGAAGATTCAGATCTCAAAGGATGGGAAAGCCCAG 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGAATTGCGGAAAACAGCCATTGAAGATTCAGATCTCAAAGGATGGGAAAGCCCAG 360

Qy    361 TAAAAATGCTTCCTAAAGACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TAAAAATGCTTCCTAAAGACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCA 420

Qy    421 AAAACAAAAGGGTAAAGAAGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAAACAAAAGGGTAAAGAAGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGC 480

Qy    481 TAAAGGTTCTAAGTCACTAAAAGTTGAGCTCATAGAGAAGATAAATAGATTATACAAAGGA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TAAAGGTTCTAAGTCACTAAAAGTTGAGCTCATAGAGAAGATAAATAGATTATACAAAGGA 540

Qy    541 AGGGCAACCATCCATAGACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCC 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGGGCAACCATCCATAGACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCC 600

Qy    601 ATATCGTTACAAGTTGGATTTTAGTCTACAGCCTGAAACAGGACCAGCAATCTCATTGA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATATCGTTACAAGTTGGATTTTAGTCTACAGCCTGAAACAGGACCAGCAATCTCATTGA 660

Qy    661 TCCGATACATACAAGTTCAAAGCCTTCACAAATACAGCAACAGCCACAGAGGATGTCAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TCCGATACATACAAGTTCAAAGCCTTCACAAATACAGCAACAGCCACAGNAGAGGATGTCAA 720

Qy    721 GATGAAATTTAGCAATGAGGTTTCCGATTTGCTTCAGCTTGTGATGAATTCACGTACCAA 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GATGAAATTTAGCAATGAGGTTTCCGATTTGCTTCAGCTTGTGATGAATTCACGTACCAA 780

Qy    781 TGGCACTATTCATTTGGAGTCAAAGACAACAACCCATGGGAAAATTGTTGGCATCAAAGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGGCACTATTCATTTGGAGTCAAAGACAACAACCCATGGGAAAATTGTTGGCATCAAAGT 840

Qy    841 CACCAATGATACCAAGGAGAGCCCTCATTCAATTCAATCTGATGATAAACAAGTATTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CACCAATGATACCAAGGAGAGCCCTCATTCAATTCAATCTGATGATAAACAAGTATTT 900

Qy    901 TGAAGACCATCAAGTCCAACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTGTGGAAGT 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TGAAGACCATCAAGTCCAACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTGTGGAAGT 960

Qy    961 TTTACTGCCAAATAGTACTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACA 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTTACTGCCAAATAGTACTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACA 1020

Qy   1021 GTTCTCTGAATGCCAATATGATTACTTCCGATTAAAATGCAAAATTACAACACAAAAT 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GTTCTCTGAATGCCAATATGATTACTTCCGATTAAAATGCAAAATTACAACACAAAAT 1080

Qy   1081 ATGGGAACAAAGTAAAAAATTCTCACTATTGTGCGAGATGGACCAGCTCTAAGGACAT 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ATGGGAACAAAGTAAAAAATTCTCACTATTGTGCGAGATGGACCAGCTCTAAGGACAT 1140

Qy   1141 TACGAAAAATAAAGTTGATTTCAGAGCATTTAAAGCAGATTTAAAACACTGGCAGAGTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TACGAAAAATAAAGTTGATTTCAGAGCATTTAAAGCAGATTTAAAACACTGGCAGAGTC 1200

Qy   1201 CAGAAAAGCAGCAGAGAAAATTCAGAGCAAAACAAATAAAAGAAGAGAGGGACC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CAGAAAAGCAGCAGAGAAAATTCAGAGCAAAACAAATAAAAGAAGAGAGGGACC 1260

Qy   1261 AAAGTTGGTTAAATTATTGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACA 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AAAGTTGGTTAAATTATTGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACA 1320

Qy   1321 GTACATTCTTGTGAACAATAAATGCCACCCAGATCAAACAAAACTTAGATTTCCTGAA 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GTACATTCTTGTGAACAATAAATGCCACCCAGATCAAACAAAACTTAGATTTCCTGAA 1380

Qy   1381 GGAAATTAAATGGTTGCTGTATTGGAGTTGATCCTGAGTCTAACATCAATGGAGTGGT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GGAAATTAAATGGTTGCTGTATTGGAGTTGATCCTGAGTCTAACATCAATGGAGTGGT 1440

Qy   1441 CAAAGCTTACAAAGGAAGAAGCCGAGTAGCAACCTTCAACCTTCCACTTTCCAAGTGTATAGAGAACA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CAAAGCTTACAAAGGAAGAAGCCGAGTAGCAACCTTCAACCTTCCACTTTCCAAGTGTATAGAGAACA 1500

Qy   1501 GAAAACCACCACCCAAATGAGACGATTTCTACTGAATCTTTACCATCAA 1549
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GAAAACCACCACCCAAATGAGACGATTTCTACTGAATCTTTACCATCAA 1549

RESULT 2
AAI24129/c
ID   AAI24129 standard; DNA; 4778 BP.
XX
AC   AAI24129;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #14062 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
```

```
PD    09-AUG-2001.
XX
PF    30-JAN-2001; 2001WO-US000670.
XX
PR    04-FEB-2000; 2000US-0180312P.
PR    26-MAY-2000; 2000US-0207456P.
PR    30-JUN-2000; 2000US-0608408.
PR    03-AUG-2000; 2000US-00632366.
PR    21-SEP-2000; 2000US-0234687P.
PR    27-SEP-2000; 2000US-0236359P.
PR    04-OCT-2000; 2000GB-00024263.
XX
PA    (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI    Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR    WPI; 2001-488901/53.
XX
PT    Human genome-derived single exon nucleic acid probes useful for analyzing
PT    gene expression in human cervical epithelial cells.
XX
PS    Claim 25; SEQ ID NO 14062; 487pp; English.
XX
CC    The present invention relates to human single exon nucleic acid probes
CC    (SENP). The present sequence is one such probe. The SENPs are derived
CC    from human HeLa cells. The SENPs can be used to produce a single exon
CC    microarray, which can be used for measuring human gene expression in a
CC    sample derived from human cervical epithelial cells. By measuring gene
CC    expression, the probes are therefore useful in grading and/or staging of
CC    diseases of the cervix, notably cervical cancer. Note: The sequence data
CC    for this patent did not form part of the printed specification, but was
CC    obtained in electronic format directly from WIPO at
CC    ftp.wipo.int/pub/published_pct_sequences
XX
SQ    Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;

Query Match        94.9%;    Score 1470.4;    DB 4;    Length 4778;
Best Local Similarity   99.9%;    Pred. No. 0;
Matches 1471;    Conservative   0;    Mismatches   0;    Indels   1;    Gaps   0;

Qy     78 GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG 137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4778 GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG 4719

Qy    138 ATGTAAATCAGTTGGTTAGAAAGTCATAAGAGTCGTTGATCTTGTTGATATGG 197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4718 ATGTAAATCAGTTGGTTAGAAAGTCATAAGAGTCGTTGATCTTGTTGATATGG 4659

Qy    198 AACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGG 257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4658 AACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGG 4599

Qy    258 GCATCACACATGGACCAGCTATTCAAATAGAGAACTATTCAAAGAATTGCGGAAAACAG 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4598 GCATCACACATGGACCAGCTATTCAAATAGAGAACTATTCAAAGAATTGCGGAAAACAG 4539

Qy    318 CCATTGAAGATTCGATTCAGATTCAGAAAGTCTAAGATGGGAAAGCCCAGTAAAAATGTCCTAAAG 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4538 CCATTGAAGATTCGATTCAGATTCAGAAAGTCTAAGATGGGAAAGCCCAGTAAAAATGTCCTAAAG 4479

Qy    378 ACCAAAACTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAACAAAAGGGTAAAG 437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4478 ACCAAAACTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAACAAAAGGGTAAAG 4419

Qy    438 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCAC 497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4418 AGAACGATTTCACTCTAAATCTTTACCATCAC 4359

Qy    498 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAG 557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4358 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAG 4299

Qy    558 ACCTGACATGTGTATCATCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGG 617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4298 ACCTGACATGTGTATCATCATTTGATGAATTCAGTAATCCGATACATGAATTCA 4239

Qy    618 ATTTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCA 677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4238 ATTTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCA 4179

Qy    678 AAGCCCTCACAAATACAGCAACAGCCACAGAGGATGTCAAGATGAAATTTAGCAATG 737
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4178 AAGCCCTCACAAATACAGCAACAGCCACAGAGGATGTCAAGATGAAATTTAGCAATG 4119

Qy    738 AGGTTTTCCGATTTGCTTCAGCTTGTGATGAATTCAGTACCAATGGCACTATTCATTTTG 797
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4118 AGGTTTTCCGATTTGCTTCAGCTTGTGATGAATTCAGTACCAATGGCACTATTCATTTTG 4059

Qy    798 GAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4058 GAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 3999

Qy    858 AAGCCCTCATTCAACCATTCAATCTGATGATAAACAAGTATTTTGAAGACCATCCAAGTCC 917
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3998 AAGCCCTCATTCAACCATTCAATCTGATGATAAACAAGTATTTTGAAGACCATCCAAGTCC 3939

Qy    918 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTA 977
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3938 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTTACTGCCAAATAGTA 3879

Qy    978 CTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT 1037
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3878 CTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT 3819

Qy   1038 ATGATTACTTCCAGATTAAAAATGCAAAATTACAACAACAAAATATGGGAACAAAGTAAAA 1097
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3818 ATGATTATTTCCAGATTAAAAATGCAAAATTACAACAACAAAATATGGGAACAAAGTAAAA 3759

Qy   1098 AATTCTCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAAAATAAAGTTG 1157
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3758 AATTCTCACTATTTGTGCGAGATGGGACCAGCTCTCAAGGACATTACGAAAAATAAAGTTG 3699

Qy   1158 ATTTCAGAGACATTTAAAGCCAGATTTTAAAACACTGCAGAGTCCAGAAAAGCAGCAGAAG 1217
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3698 ATTTCAGAGACATTTAAAGCCAGATTTTAAAACACTGCAGAGTCCAGAAAAGCAGCAGAAG 3639

Qy   1218 AAAAATTCAGAGCAAAAACAAATAAAAAGACAGAAAGAGAGGACCAAAGTTGGTTAAATTAT 1277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3638 AAAAATTCAGAGCAAAAACAAATAAAAAGACAGAAAGAGAGGACCAAAGTTGGTTAAATTAT 3579

Qy   1278 TGACAGGAAATCAAGATTGTTAGATAATTCATACTATGAACAGTGAACAGTACATTCTTGTTAACAA 1337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3578 TGACAGGAAATCAAGATTGTTAGATAATTCATACTATGAACAGTACATTCTTGTTAACAA 3519

Qy   1338 ATAAATGCCACCCAGATCAAACAAACAAACAACACTTAGATTTCCTGAAGGAAATAAATGGTTTG 1397
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3518 ATAAATGCCACCCAGATCAAACAAACAAACAACACTTAGATTTCCTGAAGGAAATAAATGGTTTG 3459

Qy   1398 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAACATCAATGAGAAA 1457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3458 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGAGTCTAACATCAATGAGAAA 3399

Qy   1458 GCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTGTATATGTGTAGAACAGAAACCACCAAATG 1517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3398 GCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTGTATATGTGTAGAACAGAAACCACCAAATG 3339

Qy   1518 AGACGATTTCACTCTAAATCTTTACCATCAA 1549
          ||||||||||||||||||||||||||||||
Db   3338 AGACGATTTCACTCTAAATCTTTACCATCAA 3307


RESULT 3
ABA69252/c
ID   ABA69252   standard; DNA; 4778 BP.
XX
AC   ABA69252;
```

XX
DT   01-FEB-2002 (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #17557.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PR   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-0060840B.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 4; SEQ ID NO 17557; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;

Query Match   94.9%; Score 1470.4; DB 4; Length 4778;
Best Local Similarity   99.9%; Pred. No. 0;
Matches 1471; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

```
Qy     78 GTATCAGAATGGCAAAGCAACTCAACCTTCCAGAAAATACAGATGATTGGACAAAGAGG 137
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4778 GTATCAGAATGGCAAAGCAACTCAACCTTCCAGAAAATACAGATGATTGGACAAAGAGG 4719

Qy    138 ATGTAAATCAGTGGTTAGAAGTCATAAGAGTCAGGGAAATTTTGACTG 197
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4718 ATGTAAATCAGTGGTTAGAAGTCATAAGAGTCAGGGAAATTTTGACTG 4659

Qy    198 AACAAGACGTGAATGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGG 257
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4658 AACAAGACGTGAATGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGG 4599

Qy    258 GCATCACACAATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAG 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4598 GCATCACACAATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAG 4539

Qy    318 CCATTGAAGATTCGATTCAGACAATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAG 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4538 CCATTGAAGATTCGATTCAGACAATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAG 4479

Qy    378 ACCAAACTGTGTCTCAAAAGGAAGTAGAACTTCAAAGCAAAAACAACAAAAGGGTAAAG 437
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4478 ACCAAACTGTGTCTCAAAAGGAAGTAGAACTTCAAAGCAAAAACAACAAAAGGGTAAAG 4419
```

```
Qy    438 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAGGTTCTAAGTCAC 497
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4418 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAGGTTCTAAGTCAC 4359

Qy    498 TAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAG 557
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4358 TAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAG 4299

Qy    558 ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCATCGTTACAAGTTGG 617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4298 ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCATCGTTACAAGTTGG 4239

Qy    618 ATTTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCA 677
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4238 ATTTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCA 4179

Qy    678 AAGCCTTCACAAATACAGCAACAGCCACAGAGAGATGTCAAGATGAAATTTAGCAATG 737
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4178 AAGCCTTCACAAATACAGCAACAGCCACAGAGAGATGTCAAGATGAAATTTAGCAATG 4119

Qy    738 AGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTG 797
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4118 AGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTG 4059

Qy    798 GAGTCAAAGACAAACCCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 857
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4058 GAGTCAAAGACAAACCCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 3999

Qy    858 AAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCC 917
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3998 AAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCC 3939

Qy    918 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTTACTGCCAAATAGTA 977
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3938 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTTACTGCCAAATAGTA 3879

Qy    978 CTTCTATCTGACAGATTGTTATTGAAGTGGACATTATTCCCACAGTTCTCTGAATGCCAAT 1037
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3878 CTTCTATCTGACAGATTGTTATTGAAGTGGACATTATTCCCACAGTTCTCTGAATGCCAAT 3819

Qy   1038 ATGATTACTTCCAGATTAAAATGCAAAATTACAACAACAAATATGGGAACAAAGTAAAA 1097
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3818 ATGATTATTCCAGATTAAAATGCAAAATTACAACAACAAATATGGGAACAAAGTAAAA 3759

Qy   1098 AATTCTCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAATTAAAGTTG 1157
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3758 AATTCTCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAATTAAAGTTG 3699

Qy   1158 ATTTCAGAGCATTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAACTGGACAAGAAG 1217
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3698 ATTTCAGAGCATTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAACTGGACAAGAAG 3639

Qy   1218 AAAAATTCAGAGCAAAACAAATACAGAGTTGGTTAAATTAT 1277
          |||||||||||||||||||||||||||||||||||||||||
Db   3638 AAAAATTCAGAGCAAAACAAATACAGAGTTGGTTAAATTAT 3579

Qy   1278 TGACAGGAAATCAAGATTTGTTAGTATAATTCATACTATGAACAGTACATTCTTGTAACAA 1337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3578 TGACAGGAAATCAAGATTTGTTAGTATAATTCATACTATGAACAGTACATTCTTGTAACAA 3519

Qy   1338 ATAAATGCCACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAAATTAAATGGTTTG 1397
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3518 ATAAATGCCACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAAATTAAATGGTTTG 3459

Qy   1398 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAAA 1457
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3458 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAAA 3399

Qy   1458 GCCGAGTAGCAAACTTCACTTTCCAAGTGTATATGTAGAACAAAAACACCACCCAAATG 1517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3398 GCCGAGTAGCAAACTTCACTTTCCAAGTGTATATGTAGAACAAAAACACCACCCAAATG 3339

Qy   1518 AGACGATTTCTACTCTAAATCTTTACCATCAA 1549
```

```
Db     3338 AGACGATTTCTACTCTCTAAATCTTTACCATCAA 3307

RESULT 4
AAI49425/c
ID    AAI49425 standard; DNA; 4778 BP.
XX
AC    AAI49425;
XX
DT    17-OCT-2001 (first entry)
XX
DE    Probe #18111 used to measure gene expression in human placenta sample.
XX
KW    Probe; microarray; human; placenta; antenatal diagnosis;
KW    genetic disorder; ss.
XX
OS    Homo sapiens.
XX
PN    WO200157272-A2.
XX
PD    09-AUG-2001.
XX
PF    30-JAN-2001; 2001WO-US000663.
XX
PR    04-FEB-2000; 2000US-0180312P.
PR    26-MAY-2000; 2000US-0207456P.
PR    30-JUN-2000; 2000US-0060408.
PR    03-AUG-2000; 2000US-0633256.
PR    21-SEP-2000; 2000US-0234687P.
PR    27-SEP-2000; 2000US-0236359P.
PR    04-OCT-2000; 2000GB-00024263.
XX
PA    (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI    Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR    WPI; 2001-488897/53.
XX
PT    Human genome-derived single exon nucleic acid probes useful for analyzing
PT    gene expression in human placenta.
XX
PS    Claim 25; SEQ ID NO 18111; 654pp; English.
XX
CC    The present invention relates to single exon nucleic acid probes (SENP).
CC    The present sequence is one such probe. The probes are useful for
CC    producing a microarray for predicting, measuring and displaying gene
CC    expression in samples derived from human placenta. The probes are useful
CC    for antenatal diagnosis of human genetic disorders
XX
SQ    Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;

Query Match    94.9%;  Score 1470.4;  DB 4; Length 4778;
Best Local Similarity  99.9%;  Pred. No. 0;
Matches 1471;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps    0;
```

```
Qy     78 GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAAATACAGATGATTGGACAAAAGAGG 137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4778 GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAAATACAGATGATTGGACAAAAGAGG 4719

Qy    138 ATGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAGTCATAAGATTGACCAAAAGT 197
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   4718 ATGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAAATTTGACTG 4659

Qy    198 AACAAGACGTGAATGGACCAGCTATTCAGAAGAAGAACTATTCAAAGAAGAACTATTCAAA 257
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   4658 AACAAGACGTGAATGGACCAGCTATTCAGAAGAAGAACTATTCAAAGAAGAACATCTTGTT 4599

Qy    258 GCATCACACATGGCACCAGCTATTCAGACATCAATCAAACAAAACAGAAGAATTGCGAAAA 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4598 GCATCACACATGGCACCAGCTATTCAAAGAAGAACTATTCAAAGAATTAAAGAAGAACAG 4539

Qy    318 CCATTGAAGAATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAAAAGTCTCTAAAG 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4538 CCATTGAAGAATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAAAACAAAGGTAAAG 4479

Qy    378 ACCAAACTGTGTCTCAAAAGGACGTAGAGAACTTCAAAGCAAAACAAAGGGTAAAG 437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4478 ACCAAACTGTGTCTCAAAAGGACGTAGAGAACTTCAAAGCAAAACAAAGGGTAAAG 4419

Qy    438 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTGAGTACAACTGCTAAGTTCTAAGTCAC 497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4418 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTGAGTACAACTGCTAAGGTTCTAAGTCAC 4359

Qy    498 TAAAAGTTGAGCTCATAGAAGATAAAATAGAGATTATACAAAGGAGAAGGCAACCATCCATAG 557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4358 TAAAAGTTGAGCTCATAGAAGATAAAATAGAGATTATACAAAGGAGAAGGCAACCATCCATAG 4299

Qy    558 ACCTGACACATGTGATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGG 617
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4298 ACCTGACACATGTGATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGG 4239

Qy    618 ATTTTAGTCTACAGCCTGAAACAGACAGGACCAATCTCATTGATCCGATACATGAATTCA 677
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4238 ATTTTAGTCTACAGCCTGAAACAGACAGGACCAATCTCATTGATCCGATACATGAATTCA 4179

Qy    678 AAGCCTTCACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAATTTAGCAATG 737
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4178 AAGCCCTTCACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAATTTAGCAATG 4119

Qy    738 AGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGCACTATTCATTTTG 797
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4118 AGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGCACTATTCATTTTG 4059

Qy    798 GAGTCAAAGACAAACCCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 857
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4058 GAGTCAAAGACAAACCCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 3999

Qy    858 AAGCCCTCATTAACCATTTCAATCTGATGATGAAACAAGTATTTGAAGACCATCAAGTCC 917
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3998 AAGCCCTCATTAACCATTTCAATCTGATGATGAAACAAGTATTTGAAGACCATCAAGTCC 3939

Qy    918 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTA 977
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3938 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTA 3879

Qy    978 CTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT 1037
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3878 CTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT 3819

Qy   1038 ATGATTACTTCCAGATTAAATAAATGCAAAATTACAACACAAAATATGGGAACAAAGTAAAA 1097
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3818 ATGATTATTTCCAGATTAAATAAATGCAAAATTACAACAACAAAATATGGGAACAAAGTAAA 3759

Qy   1098 AATTCTCACTATTTGTGCGAGATGGGACCAGCTCAAGGACCTCAAGGAACATTACGAAAAATAAAGTTG 1157
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3758 AATTCTCACTATTTGTGCGAGATGGGACCAGCTCAAGGACCTCAAGGAACATTACGAAAAAATAAAGTTG 3699

Qy   1158 ATTTCAGAGCATTTAAAGCAGATTTTAAAGCACTGGCAGAGTCCAGAAAAGCAGCAGAAGG 1217
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3698 ATTTCAGAGCATTTAAAGCAGATTTTAAAGCACTGGCAGAGTCCAGAAAAGCAGCAGAAG 3639

Qy   1218 AAAAAATTCAGAGCAAAACAAATAAAAAGAAGAAACACAGGGAAATTTGACTG 1277
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3638 AAAAAATTCAGAGCAAAACAAATAAAAAGAAGAAACACAGGGAAATTTGACTG 3579

Qy   1278 TGACAGGAAATCAAGATTTGTTAGATAATTCATATGAACAGTACATTCTTGTAACAA 1337
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3578 TGACAGGAAATCAAGATTTGTTAGATAATTCATATGAACAGTACATTCTTGTAACAA 3519

Qy   1338 ATAAATGCCACCCAGATCAAACAAAACAAACACTTAGATTTCCTGAAGGAAATTAAAATGTTTTG 1397
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3518 ATAAATGCCACCCAGATCAAACAAAACAAACACTTAGATTTCCTGAAGGAAATTAAAATGTTTTG 3459

Qy   1398 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCTAACATCAATGAGAAGAAA 1457
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Query Match           94.9%;  Score 1470.4;  DB 4;  Length 4778;
Best Local Similarity  99.9%;
Matches 1471;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy         78 GTATCAGAATGGCAAAGCAACTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4778 GTATCAGAATGGCAAAGCAACTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG 4719

Qy        138 ATGTAAATCAGTGGTTAGAAAGTCATAAGATCATTGACCAAAAACACAGGGAAATTTGACTG 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4718 ATGTAAATCAGTGGTTAGAAAGTCATAAGATCATTGACCAAAAACACAGGGAAATTTGACTG 4659

Qy        198 AACAAGACGTGAATGGAGCAGTCTTGAGGTGGTTAAAAAAGAACATCTTGTTGATATGG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4658 AACAAGACGTGAATGGAGCAGTCTTGAGGTGGTTAAAAAAGAACATCTTGTTGATATGG 4599

Qy        258 GCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAG 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4598 GCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAG 4539

Qy        318 CCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAG 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4538 CCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAG 4479

Qy        378 ACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAACAAAAGGGTAAAG 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4478 ACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAACAAAAGGGTAAAG 4419

Qy        438 AGAACCCAGATAATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCAC 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4418 AGAACCCAGATAATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCAC 4359

Qy        498 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAG 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4358 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAG 4299

Qy        558 ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGG 617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4298 ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGG 4239

Qy        618 ATTTTAGTCTACAGCCTGAAACAGGAACAAGGACCAGGCAATCTCATTGATCCGATACATGAATTCA 677
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4238 ATTTTAGTCTACAGCCTGAAACAGGAACAAGGACCAGGCAATCTCATTGATCCGATACATGAATTCA 4179

Qy        678 AAGCCTTCACAAATACAGCAAGAGGAGATGTCAAGATGAAATTTAGCAATG 737
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4178 AAGCCTTCACAAATACAGCAAGAGGAGATGTCAAGATGAAATTTAGCAATG 4119

Qy        738 AGGTTTTCCGATTTGCTTCAGCTTGTGATGAATTCACCTGACCATTCATTTTG 797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4118 AGGTTTTCCGATTTGCTTCAGCTTGTGATGAATTCACCTGACCATTCATTTTG 4059

Qy        798 GAGTCAAAGCAACAAACCCATGGGAAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 857
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4058 GAGTCAAAGCAACAAACCCATGGGAAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 3999

Qy        858 AAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTGAAGACCATCAAGTCC 917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3998 AAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTGAAGACCATCAAGTCC 3939

Qy        918 AACAAGCAAAGAAGTGCATTGAGAGCCAAGATTGTGGAAGTTTACTGCCAAATAGTA 977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3938 AACAAGCAAAGAAGTGCATTGAGAGCCAAGATTGTGGAAGTTTACTGCCAAATAGTA 3879

Qy        978 CTCTATCTGACAGATTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT 1037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3878 CTCTATCTGACAGATTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT 3819

Qy       1038 ATGATTACTTCCAGATAAAATGCAAAATACACACAACAAAATATGGAACAAAGTAAAA 1097
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3818 ATGATTATTTCCAGATAAAATGCAAAATACACACAACAAAATATGGAACAAAGTAAAA 3759

Qy       1098 AATTCTCACTATTTGTGCGAGATGGGACCCAGCTTCAAGGACATTACGAAAAATAAAGTTG 1157
```

RESULT 5

Db	3458 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCTTACAAAGAAA 3399
Qy	1458 GCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAACACCACCAAATG 1517
Db	3398 GCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAACCACCACCAAATG 3339
Qy	1518 AGACGATTTCTACTCTAAATCTTTACCATCAA 1549
Db	3338 AGACGATTTCTACTCTAAATCTTTACCATCAA 3307

ABA51244/c
ID ABA51244 standard; DNA; 4778 BP.
XX
AC ABA51244;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #9939.
XX
KW Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000662.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX
PS Claim 4; SEQ ID NO 9939; 327pp + Sequence Listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a single exon
CC nucleic acid probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;

CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;

Query Match   94.9%; Score 1470.4; DB 4; Length 4778;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   78 GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG 137
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4778 GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG 4719

Qy   138 ATGTAAATCAGTGGTTAGAAAGTCATAAGAATTGACCAAAAACACAGGGAAATTTGACTG 197
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4718 ATGTAAATCAGTGGTTAGAAAGTCATAAGAATTGACCAAAAACACAGGGAAATTTGACTG 4659

Qy   198 AACAAGACGTGAATGGACGAGCAGTCTTGAAGTGGTAAAAAAAGAACATCTTGTGATATGG 257
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4658 AACAAGACGTGAATGGACGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTTGTGATATGG 4599

Qy   258 GCATCACACATGGACCAGCTATTCAAATGACTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAG 317
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4598 GCATCACACATGGACCAGCTATTCAAATGACTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAG 4539

Qy   318 CCATTGAAGATTCGATTCAGATCATCAAGATGGAAGCCCAGTAAAAATGCTCCTAAAG 377
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4538 CCATTGAAGATTCGATTCAGATCATCAAGATGGAAGCCCAGTAAAAATGCTCCTAAAG 4479

Qy   378 ACCAAACTGTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAAAACAAAAGGTAAAG 437
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4478 ACCAAACTGTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAAAACAAAAGGTAAAG 4419

Qy   438 AGAACCCAGATATGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCAC 497
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4418 AGAACCCAGATATGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCAC 4359

Qy   498 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAG 557
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4358 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAG 4299

Qy   558 ACCTGACATGTGTATCATATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGG 617
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4298 ACCTGACATGTGTATCATATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGG 4239

Qy   618 ATTTTAGTTCACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCA 677
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4238 ATTTTAGTTCACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCA 4179

Qy   678 AAGCCTTCACAAATACAGACAACAGCCACAGAAGGATGCAAGATGAAATTAGCAATG 737
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4178 AAGCCTTCACAAATACAGACAACAGCCACAGAAGGATGCAAGATGAAATTAGCAATG 4119

Qy   738 AGGTTTCCGATTGCTTGCTTCAGCTTGTATGATGAATTCACGTACCAATGGCACTATTCATTTTG 797
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4118 AGGTTTCCGATTGCTTGCTTCAGCTTGTATGATGAATTCACGTACCAATGGCACTATTCATTTTG 4059

Qy   798 GAGTCAAAGACAAACCCCATGGGAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 857
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4058 GAGTCAAAGACAAACCCCATGGGAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 3999

Qy   858 AAGCCCTCATTCAACTGATGATAAACAAGTATTTGAAGACCATCAAGTCC 917
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3998 AAGCCCTCATTCAATCTGATGATAAACAAGTATTTGAAGACCATCAAGTCC 3939

Db   3758 AATTCTCACTATTTGTGCGACGATGGGACCAGCTCTCAAGGACATTACGACGAAAAATAAAGTTG 3699
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   1158 ATTTCAGAGCAGCATTTAAAGCAGATTTAAAACACTGGCAGAGTCCAGATAGCAGCAGAAG 1217
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3698 ATTTCAGAGCAGCATTTAAAGCAGATTTAAAACACTGGCAGAGTCCAGATAGCAGCAGAAG 3639

Qy   1218 AAAAATTCAGACAGCAAAAACAAATAAAAGAAGAAAGAGAGGGACCAAAGTTGGTTAAATTAT 1277
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3638 AAAAATTCAGACAGCAAAAACAAATAAAAGAAGAAAGAGAGGGACCAAAGTTGGTTAAATTAT 3579

Qy   1278 TGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTAACAA 1337
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3578 TGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTAACA 3519

Qy   1338 ATAAATGCCACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAAATTAAATGGTTTG 1397
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3518 ATAAATGCCACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAAATTAAATGGTTTG 3459

Qy   1398 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCTTACAAAGAAA 1457
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3458 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCTTACAAAGAAA 3399

Qy   1458 GCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAAACCACCACCAAATG 1517
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3398 GCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAAACCACCACCAAATG 3339

Qy   1518 AGACGATTTCACTCTAAATCTTTACCATCAA 1549
                 |||||||||||||||||||||||||||
Db   3338 AGACGATTTCACTCTAAATCTTTACCATCAA 3307

RESULT 6
ABA36171/c
ID ABA36171 standard; DNA; 4778 BP.
XX
AC ABA36171;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #14637 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000666.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-00633266.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.
XX
PS Claim 4; SEQ ID NO 14637; 530pp; English.
XX

```
PR
XX   04-OCT-2000; 2000GB-00024263.
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
DR
XX   WPI; 2001-488900/53.
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human bone marrow.
PS
XX
     Example 4; SEQ ID NO 17911; 658pp + Sequence Listing; English.
CC
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention
XX
SQ   Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;

Query Match              94.9%;  Score 1470.4;  DB 4;  Length 4778;
Best Local Similarity    99.9%;  Pred. No. 0;
Matches 1471;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     78 GTATCAGAATGGCAAAGCAACTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG 137
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4778 GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG 4719

Qy    138 ATGTAAATCAGTGGTTAGAAAGTCATAAGAGTCATAAGAGTCATAAGAGTTTGACTG 197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4718 ATGTAAATCAGTGGTTAGAAAGTCATAAGAGTCATAAGAGTCATAAGAGTTTGACTG 4659

Qy    198 AACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTTGTTGATATGG 257
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4658 AACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTTGTTGATATGG 4599

Qy    258 GCATCACACATGGACCACCAGCTATTCAAATAGAAGAACTATTCAAAGAATGGGAAAACAG 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4598 GCATCACACATGGACCACCAGCTATTCAAATAGAAGAACTATTCAAAGAATGGGAAAACAG 4539

Qy    318 CCATTGAAGATTCGATTCAGACATCTAAGATGGAAGCCCAGTAAAAATGCTCCTAAAG 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4538 CCATTGAAGATTCGATTCAGACATCTAAGATGGAAGCCCAGTAAAAATGCTCCTAAAG 4479

Qy    378 ACCAAACTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAACAAAAGGGTAAAG 437
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4478 ACCAAACTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAACAAAAGGGTAAAG 4419

Qy    438 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAGGTTCTAAGTCAC 497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4418 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAGGTTCTAAGTCAC 4359

Qy    498 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAAGCAACCATCCATAG 557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4358 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAAGCAACCATCCATAG 4299

Qy    558 ACCTGACATGTGTATCATCATTTGATGAATTCATTTGATGAATTCATTATCGTTACAAGTTGG 617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4298 ACCTGACATGTGTATCATCATTTGATGAATTCATTTGATGAATTCATTATCGTTACAAGTTGG 4239

Qy    618 ATTTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCA 677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4238 ATTTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCA 4179

Qy    678 AAGCCTTCACAAATACAGCAACAGCCCACAGAGAGGATGTCAAGATGAAATTTAGCAATG 737
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4178 AAGCCTTCACAAATACAGCAACAGCCCACAGAGAGGATGTCAAGATGAAATTTAGCAATG 4119

Qy    738 AGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATTCATTTTG 797
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4118 AGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATTCATTTTG 4059

Qy    918 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGAGATTTGTGGAAGTTTTACTGCCAAATAGTA 977
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3938 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGAGATTTGTGGAAGTTTTACTGCCAAATAGTA 3879

Qy    978 CTCTATCTCGACAGATTGGTTATTGGAAGTGGACATTATTCCACAGTTCTCGAATGCCAAT 1037
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3878 CTCTATCTCGACAGATTGGTTATTGGAAGTGGACATTATTCCACAGTTCTCGAATGCCAAT 3819

Qy   1038 ATGATTACTTCCAGATTAAAATGCAAAATTACAACAACAAATATGGAACAAAGTAAAA 1097
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3818 ATGATTATTTCCAGATTAAAATGCAAAATTACAACAACAAATATGGAACAAAGTAAAA 3759

Qy   1098 AATTCTCACTATTGTGCGAGATGGACCAGCTCTTAAGGACATTACGAAAAATAAAGTTG 1157
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3758 AATTCTCACTATTGTGCGAGATGGACCAGCTCTTAAGGACATTACGAAAAATAAAGTTG 3699

Qy   1158 ATTTCAGAGCAGATTTAAAGCAGATTTAAAACACTGGCAGAGTCCAGAAAAGCAGCAGAAG 1217
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3698 ATTTCAGAGCAGATTTAAAGCAGATTTAAAACACTGGCAGAGTCCAGAAAAGCAGCAGAAG 3639

Qy   1218 AAAAATTCAGAGCAGCAAAACAAATAAAAAAGAAAGAGAGGACCAAAGTTGGTTAAATTAT 1277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3638 AAAAATTCAGAGCAGCAAAACAAATAAAAAAGAAAGAGAGGACCAAAGTTGGTTAAATTAT 3579

Qy   1278 TGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTGTAACAA 1337
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3578 TGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTGTAACAA 3519

Qy   1338 ATAAATGCCACCCAGATCAAAACAAAACACTTAGATTCCTGAAGGAAATTAAATGGTTTG 1397
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3518 ATAAATGCCACCCAGATCAAAACAAAACACTTAGATTCCTGAAGGAAATTAAATGGTTTG 3459

Qy   1398 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGAGTCGTCAAAGCTTACAAAGAAA 1457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3458 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGAGTCGTCAAAGCTTACAAAGAAA 3399

Qy   1458 GCCGAGTAGCAAACCTTCACTTTCCAAGTGTATATGTGAACAGCAGAAACCACACCAAATG 1517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3398 GCCGAGTAGCAAACCTTCACTTTCCAAGTGTATATGTGAACAGCAGAAACCACCACCAAATG 3339

Qy   1518 AGACGATTCTACTCTAAATCTTTACCATCAA 1549
          |||||||||||||||||||||||||||||||
Db   3338 AGACGATTCTACTCTAAATCTTTACCATCAA 3307
```

```
RESULT 7
AAK43354/c
ID   AAK43354  standard; DNA; 4778 BP.
XX
AC   AAK43354;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 17911.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000668.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-0063366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
```

```
XX
PF   30-JAN-2001; 2001WO-US000667.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains.
XX
PS   Example 4; SEQ ID NO 17536; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention
XX
SQ   Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;

Query Match             94.9%;  Score 1470.4;  DB 4;  Length 4778;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 1471;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  0;

Qy     78 GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGAG 137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4778 GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGAG 4719

Qy    138 ATGTAAATCAGTGGTTAGAAAGTCATAAGTAGAAGATTCATAAGAATACACAGGGAAATTTTGACTG 197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4718 ATGTAAATCAGTGGTTAGAAAGTCATAAGTAGAAGATTCATAAGAATACACAGGGAAATTTTGACTG 4659

Qy    198 AACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTTGTTGATATGG 257
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4658 AACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTTGTTGATATGG 4599

Qy    258 GCATCACACATGACCAGCTATTCAAATCAAATAGAAGAACTATTCAAAGAATTGCGAAAACAG 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4598 GCATCACACATGACCAGCTATTCAAATCAAATAGAAGAACTATTCAAAGAATTGCGAAAACAG 4539

Qy    318 CCATTGAAGATTCGATTCAGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAG 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4538 CCATTGAAGATTCGATTCAGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAG 4479

Qy    378 ACCAAACTGTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAAGCAAAAGGGTAAAG 437
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4478 ACCAAACTGTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAAGCAAAAGGGTAAAG 4419

Qy    438 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTGAGTACAACTGCTAAAGGTTCTAAGTCAC 497
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4418 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTGAGTACAACTGCTAAAGGTTCTAAGTCAC 4359

Qy    498 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAAGGAAAGGCAACCATCCATAG 557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4358 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAAGGAAAGGCAACCATCCATAG 4299

Qy    558 ACCTGACATGTGTAATCATATATCG 617
          |||||||||||||||||||||||
Db   4298 ACCTGACATGTGTAATCATATATCG 4239

Qy    618 ATTTTAGTTCACAGCCTGAAACAGCAGGCAATCTCATTGATCCGATACATGAATTCA 677
```

```
Qy    798 GAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4058 GAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 3999

Qy    858 AAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTGAAGACCATCAAGTCC 917
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3998 AAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTGAAGACCATCAAGTCC 3939

Qy    918 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTA 977
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3938 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTA 3879

Qy    978 CTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT 1037
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3878 CTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT 3819

Qy   1038 ATGATTACTTCCAGATTAAAATGCAAAATTACAACAACAAATATGGGAACAAAGTAAAA 1097
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   3818 ATGATTATTTCCAGATTAAAATGCAAAATTACAACAACAAATATGGGAACAAAGTAAAA 3759

Qy   1098 AATTCCACTATTTGTGCGAGATGGGACCGCTCTAAGGACATTACGAGAAAAATAAAGTTG 1157
          ||||||||||||||||||||||||||||  |||||||||||||||||| ||||||||||
Db   3758 AATTCCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAGAAAATAAAGTTG 3699

Qy   1158 ATTTCAGAGCATTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAAAGCAGCAGAAG 1217
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3698 ATTTCAGAGCATTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAAAGCAGCAGAAG 3639

Qy   1218 AAAAATTCAGAGCAGCAAAAACAAATAAAAAAGAAGAGAAGGAGGGACCAAAGTTGGTTAAATTAT 1277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3638 AAAAATTCAGAGCAGCAAAAACAAATAAAAAAGAAGAGAAGGAGGGACCAAAGTTGGTTAAATTAT 3579

Qy   1278 TGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTAACAA 1337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3578 TGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTAACAA 3519

Qy   1338 ATAAATGCCACCGATCAAACAAACACTTAGATTCCTGAAGGAAATTAAATGGTTTG 1397
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3518 ATAAATGCCACCGATCAAACAAACACTTAGATTCCTGAAGGAAATTAAATGGTTTG 3459

Qy   1398 CTGTATTGGAGTTTGATCCTGAGTCTTAACATCAATGAGTGGTCAAAGCTTACAAAGAAA 1457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3458 CTGTATTGGAGTTTGATCCTGAGTCTTAACATCAATGAGTGGTCAAAGCTTACAAAGAAA 3399

Qy   1458 GCCGAGTAGCAAACCTTCACTTTCCAAGTGTATATGTAGAACAGAAAACCACACCCAAATG 1517
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3398 GCCGAGTAGCAAACCTTCACTTTCCAAGTGTATATGTAGAACAGAAAACCACACCCAAATG 3339

Qy   1518 AGACGATTTCTACTCTAAATCTTTACCATCAA 1549
          |||||||||||||||||||||||||||||||
Db   3338 AGACGATTTCTACTCTAAATCTTTACCATCAA 3307
```

```
RESULT 8
AAK17545/c
ID   AAK17545 standard; DNA; 4778 BP.
XX
AC   AAK17545;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 17536.
XX
KW   Human; brain expressed exon; gene expression analysis; probe; microarray;
KW   Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
```

```
XX
KW   Human; single exon nucleic acid probe; liver; cirrhosis;
KW   hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW   coronary heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157273-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000664.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-06S2366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488898/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human adult liver.
XX
PS   Claim 4; SEQ ID NO 17966; 658pp; English.
XX
CC   The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC   measuring human gene expression in a sample derived from human adult
CC   liver, comprising one of 13109 defined nucleotide sequences given in the
CC   specification (or complements/ fragments). The probe hybridises at high
CC   stringency to a nucleic acid molecule expressed in the human adult liver.
CC   (I) may be used for predicting, measuring and displaying gene expression
CC   in samples derived from human adult liver. The genes identified may be
CC   involved in genetic liver diseases such as cirrhosis,
CC   hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC   associated with coronary heart disease. ABS25011-ABS51005 represent human
CC   liver single exon nucleic acid probes of the invention. Note: The
CC   sequence information for this patent does not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;

Query Match             94.9%;  Score 1470.4;  DB 4;  Length 4778;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 1471;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    0;

Qy     78 GTATCAGAATGGCAAAGCAACTTCAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG    137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4778 GTATCAGAATGGCAAAGCAACTTCAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG   4719

Qy    138 ATGTAAATCAGTGGTTAGAAAGTCATAAGAGTTCATAAGAGTTCAAAGGGAAATTTTGACTG    197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4718 ATGTAAATCAGTGGTTAGAAAGTCATAAGAGTTCATAAGAGTTCAAAGGGAAATTTTGACTG   4659

Qy    198 AACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGG    257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4658 AACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGG   4599

Qy    258 GCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAACAG    317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4598 GCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAACAG   4539

Qy    318 CCATTGAAGAATTCGATTCAGCAATCATCTCAAGATGGGAAGCCCAGTAAAAATGTCCTAAAG    377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4538 CCATTGAAGAATTCGATTCAGCAATCATCTCAAGATGGGAAGCCCAGTAAAAATGTCCTAAAG   4479

Db   4238 ATTTTAGTGCTACAGCCTGAAACAGGACCAATCTCATTGTCATTGATCCGATACATGAATTCA   4179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    678 AAGCCTTCACAAATACAGCAACAGCCCAGAGGATGTCAAGATGAAATTTAGCAATG    737
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4178 AAGCCTTCACAAATACAGCAACAGCCCAGAGGATGTCAAGATGAAATTTAGCAATG   4119

Qy    738 AGGTTTCCGATTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTG    797
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4118 AGGTTTCCGATTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTG   4059

Qy    798 GAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG    857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4058 GAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG   3999

Qy    858 AAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCC    917
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3998 AAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCC   3939

Qy    918 AACAAGCAAAGAAGAGTGCATTCGAGAGCCAAGATTGTTGTGGAAGTTTACTGCCAAATAGTA    977
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3938 AACAAGCAAAGAAGAGTGCATTCGAGAGCCAAGATTGTTGTGGAAGTTTACTGCCAAATAGTA   3879

Qy    978 CTCTATCTGACAGATTTGTTATTGTAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT   1037
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3878 CTCTATCTGACAGATTTGTTATTGTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT   3819

Qy   1038 ATGATTACTTCCAGATTAAAATGCAAAATTACACACAACAAAAATATGGGAACAAAGTAAAA   1097
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3818 ATGATTATTTCCAGATTAAAATGCAAAATTACACACAACAACAAAAATATGGGAACAAAGTAAAA   3759

Qy   1098 AATTCTCACTATTGTGCGAGATGGACCCAGCTCTAAGGACATTACGAAAAATAAAGTG   1157
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3758 AATTCTCACTATTGTGCGAGATGGACCCAGCTCTAAGGACATTACGAAAAATAAAGTTG   3699

Qy   1158 ATTTCAGAGGACATTTAAAAGCAGATTTTAAAACACTGGCAGTGTCCAGAGAAAAGCAGCAGAAG   1217
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3698 ATTTCAGAGGACATTTAAAAGCAGATTTTAAAACACTGGCAGTGTCCAGAGAAAAGCAGCAGAAG   3639

Qy   1218 AAAAATTCAGACAGACAAAAACAAAATAAAAAAAAGAAAGAGAGGGACCAAAGTTGGTAAAATAT   1277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3638 AAAAATTCAGACAGCAAAAACAAAATAAAAAAAAGAAAGAGAGGGACCAAAGTTGGTAAAATAT   3579

Qy   1278 TGACAGGAAATCAAGATTCTAGATAATTCATCATACTACTATGAACAGTACATTCTTGTGTAACAA   1337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3578 TGACAGGAAATCAAGATTTGTTAGATAATTCATCATACTACTATGAACAGTACATTCTTGTGTAACAA   3519

Qy   1338 ATAAATGCCACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAAATTAAAATGGTTTG   1397
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3518 ATAAATGCCCCCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAAATTAAAATGGTTTG   3459

Qy   1398 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCTTACAAAGAAA   1457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3458 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCTTACAAAGAAA   3399

Qy   1458 GCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAACCACCACCAAATG   1517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3398 GCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAACCACCACCAAATG   3339

Qy   1518 AGACGATTTCTACTCTAAATCTTTACCATCAA   1549
          ||||||||||||||||||||||||||||||||
Db   3338 AGACGATTTCTACTCTAAATCTTTACCATCAA   3307

RESULT 9
ABS42976/c
ID   ABS42976 standard; DNA; 4778 BP.
XX
AC   ABS42976;
XX
DT   25-FEB-2003 (first entry)
XX
DE   Human liver single exon probe, SEQ ID No 17966.
```

```
Db    3398  GCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAGAAACCACACCCAAATG  3339
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy    1518  AGAGCGATTTCTACTCTAAATCTTTACCATCAA  1549
            |||||||||||||||||||||||||||||||||
Db    3338  AGAGCGATTTCTACTCTAAATCTTTACCATCAA  3307


RESULT 10
AAI09703/c
ID   AAI09703 standard; DNA; 4778 BP.
XX
AC   AAI09703;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Probe #9694 used to measure gene expression in human breast sample.
XX
KW   Probe; human; breast disease; breast cancer; development disorder; 88;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US000661.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00633366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,   Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression in
PT   a human breast.
XX
PS   Claim 25; SEQ ID NO 9694; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present invention is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and non-
CC   carcinoma tumours. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;

Query Match     94.9%;  Score 1470.4;  DB 5;  Length 4778;
Best Local Similarity 99.9%;  Pred. No. 0;
Matches 1471;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  0;

Qy      78  GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG  137
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4778  GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG  4719

Qy     138  ATGTAAATCAGTTGTTGAAGTCATCAAAGTCATTAAGAATGTGACCAAAGATTTTGACTG  197
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy     378  ACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACAAAAGGGTAAAG  437
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4478  ACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACAAAAGGGTAAAG  4419

Qy     438  AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCAC  497
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4418  AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCAC  4359

Qy     498  TAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAG  557
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4358  TAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAG  4299

Qy     558  ACCTGACACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGG  617
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4298  ACCTGACACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGG  4239

Qy     618  ATTTTAGTCTACAGCCTGAAACAGGACCCAGGCCAATCTCATTGATCCGATACATGAATTCA  677
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4238  ATTTTAGTCTACAGCCTGAAACAGGACCCAGGCCAATCTCATTGATCCGATACATGAATTCA  4179

Qy     678  AAGCCTTCACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGCAATG  737
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4178  AAGCCTTCACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGCAATG  4119

Qy     738  AGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTG  797
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4118  AGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTG  4059

Qy     798  GAGTCAAAGACAAACCCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG  857
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4058  GAGTCAAAGACAAACCCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG  3999

Qy     858  AAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCC  917
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3998  AAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCC  3939

Qy     918  AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTA  977
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3938  AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTA  3879

Qy     978  CTCTATCTGACAGATTTGTATTGGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT  1037
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3878  CTCTATCTGACAGATTTGTATTGGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT  3819

Qy    1038  ATGATTACTTCCAGATTAAAATGCAAAATACAACAACAAAAAATATGGGAACAAGTAAAA  1097
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3818  ATGATTATTTCCAGATTAAAATGCAAAATACAACAACAAAAAATATGGGAACAAGTAAAA  3759

Qy    1098  AATTCTCACTATTTGTGCGAGATGGGACGACCAGCTCTAAGGACATTACGAGAAAAATAAAGTTG  1157
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3758  AATTCTCACTATTTGTGCGAGATGGGACGACCAGCTCTAAGGACATTACGAGAAAAATAAAGTTG  3699

Qy    1158  ATTTCAGAGACATTTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAAGCAGCAGAG  1217
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3698  ATTTCAGAGACATTTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAAGCAGCAGAG  3639

Qy    1218  AAAAATTCAGAGACAAAAACAAAATAAAAAAGAAAGAAGAGGGACCAAAGTTGGTTAAATTAT  1277
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3638  AAAAATTCAGAGACAAAAACAAAATAAAAAAGAAAGAAGAGGGACCAAAGTTGGTTAAATTAT  3579

Qy    1278  TGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTAACAA  1337
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3578  TGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTAACAA  3519

Qy    1338  ATAAATGCCACCCAGATCAAACAAACAAAACACTTAGATTTCCTGAAGGAAATTAAATGGTTTG  1397
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3518  ATAAATGCCACCCAGATCAAACAAACAAAACACTTAGATTTCCTGAAGGAAATTAAATGGTTTG  3459

Qy    1398  CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAAA  1457
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3458  CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAAA  3399

Qy    1458  GCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAGAAACCACACCCAAATG  1517
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy  1278 TGACAGGAAATCAGAGATTTGTTAGATAATTCATACATGAACAGTACATTCTTGTAACAA 1337
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3578 TGACAGGAAATCAGAGATTTGTTAGATAATTCATACATGAACAGTACATTCTTGTAACAA 3519

Qy  1338 ATAAATGCCACCCAGATCAAACCAAACACTTAGATTCCTGAAGGAGAATTAAATGGTTTG 1397
         |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db  3518 ATAAATGCCACCCAGATCAAACAAAACACTTAGATTCCTGAAGGAGAATTAAATGGTTTG 3459

Qy  1398 CTGTATTGGAGTTTGATCCTGAGTCAACATCAATGGAGTGGTCAAAGCTTACAAAGAAA 1457
         ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db  3458 CTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCTTACAAAGAAA 3399

Qy  1458 GCCGAGTAGCAAACCTTCACTTTCCAAGTGTATATGTAGAACAGAAAACCACCACCAAATG 1517
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3398 GCCGAGTAGCAAACCTTCACTTTCCAAGTGTATATGTAGAACAGAAAACCACCACCAAATG 3339

Qy  1518 AGACGATTTCTACTCTAAATCTTTACCATCAA 1549
         ||||||||||||||||||||||||||||||||
Db  3338 AGACGATTTCTACTCTAAATCTTTACCATCAA 3307
```

RESULT 11
AB17437/c
ID   AB17437 standard; DNA; 4778 BP.
XX
AC   AB17437;
XX
DT   19-AUG-2002 (first entry)
XX
DE   Human genome-derived single exon probe ORF from lung SEQ ID No 17428.
XX
KW   Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW   pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW   primary ciliary dyskinesis; pulmonary hypertension;
KW   hyaline membrane disease; open reading frame; ORF.
XX
OS   Homo sapiens.
XX
PN   WO200186003-A2.
XX
PD   15-NOV-2001.
XX
PF   30-JAN-2001; 2001WO-US000665.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2002-114183/15.
XX
PT   Spatially-addressable set of single exon nucleic acid probes, used to
PT   measure gene expression in human lung samples.
XX
PS   Claim 4; SEQ ID NO 17428; 634pp; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human lung comprising single exon nucleic acid probes having one of
CC   12614 nucleic acid sequences mentioned in the specification, or their

```
Db  4718 ATGTAAATCAGTGGTTAGAAGTCATAAGAATTGACCAAAACACAGGGAAATTTGACTG 4659

Qy   198 AACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAAAGAACATCTTGTTGATATGG 257
         ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  4658 AACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTTGTTGATATGG 4599

Qy   258 GCATCACACATGGACCAGCTATTCAAATAGAGAGACTATTCAAAGAATTGCGGAAAACAG 317
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4598 GCATCACACATGGACCAGCTATTCAAATAGAGAGACTATTCAAAGAATTGCGGAAAAACAG 4539

Qy   318 CCATTGAAGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAG 377
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4538 CCATTGAAGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAG 4479

Qy   378 ACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTCAAAGCAAAACAAAAGGGTAAAG 437
         ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db  4478 ACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTCAAAGCAAAACAAAACTTCAAAGAAGGGTAAAG 4419

Qy   438 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGTTCAAGTCAC 497
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4418 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGTTCAAGTCAC 4359

Qy   498 TAAAAGTTGAGCTCATAGAGAGATAAAATAGTAATATACAAAGGAAAGGCAACCATCCATAG 557
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4358 TAAAAGTTGAGCTCATAGAGAGATAAAATAGTAATATACAAAGGAAAGGCAACCATCCATAG 4299

Qy   558 ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTCGG 617
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4298 ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTCGG 4239

Qy   618 ATTTTAGTCTACAGCCTGAAACAGGACCAATCTCATTGATCCGATACATGAATTCA 677
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4238 ATTTTAGTCTACAGCCTGAAACAGGACCAATCTCATTGATCCGATACATGAATTCA 4179

Qy   678 AAGCCCTCACAAATACAGCACAGCCCACAGAAGAGGATGTCAAGATGAAATTTAGCAATG 737
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4178 AAGCCCTCACAAATACAGCACAGCCCACAGAAGAGGATGTCAAGATGAAATTTAGCAATG 4119

Qy   738 AGGTTTCCGATTTGCTTCAGCTTGTGATGAATTCACGTACCAATGGCACTATTCATTTTG 797
         || ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db  4118 AGTTTTCCGATTTGCTTCAGCTTGTGATGAATTCACGTACCAATGGCACTATTCATTTG 4059

Qy   798 GAGTCAAAGACAAACCCCATGGGAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 857
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4058 GAGTCAAAGACAAACCCCATGGGAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 3999

Qy   858 AAGCCCTCATTAACACATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCC 917
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3998 AAGCCCTCATTAACACATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCC 3939

Qy   918 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTTACTGCCAAATAGTA 977
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3938 AACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTTACTGCCAAATAGTA 3879

Qy   978 CTCTATCTGACAGATTTGTATTGAAGTGACATATTCCACAGTTCTCTGAATGCCAAT 1037
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3878 CTCTATCTGACAGATTTGTATTGAAGTGACATATTCCACAGTTCTCTGAATGCCAAT 3819

Qy  1038 ATGATTACTTCCAGATTAAAATGCAAAATACAACAACAAAATATGGAACAAAGTAAAA 1097
         |||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||
Db  3818 ATGATTATTTCCAGATTAAAATGCAAAATACAACAACAAAATATGGGAACAAAGTAAAA 3759

Qy  1098 AATTCCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAAAAATAAAGTTG 1157
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3758 AATTCCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAAAAATAAAGTTG 3699

Qy  1158 ATTTCAGAGCATTTAAAGCAGATTTTAAAACACTGCAGAGTCCAGAAAGCAGCAGAG 1217
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3698 ATTTCAGAGCATTTAAAGCAGATTTTAAAACACTGCAGAGTCCAGAAAGCAGCAGAG 3639

Qy  1218 AAAAATTCAGAGCAAAACAAATAAAAAGAAAAAGAGGACCAAAGTTGGTTAAATTAT 1277
         ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  3638 AAAAATTCAGAGCAGCAAAACAAATAAAAAGAAAAAGAGGACCAAAGTTGGTTAAATTAT 3579
```

CC   complements or the 12387 open reading frames derived from the 12614
CC   probes. Also included are a microarray comprising the novel set of probes
CC   ; the novel set of probes which hybridise at high stringency to a nucleic
CC   acid expressed in the human lung; measuring gene expression in a sample
CC   derived from human lung, comprising (a) contacting the array with a
CC   collection of detectably labeled nucleic acids derived from human lung
CC   mRNA, and (b) measuring the label detectably bound to each probe of the
CC   array; identifying exons in a eukaryotic genome, comprising (a)
CC   algorithmically predicting at least one exon from genomic sequences of
CC   the eukaryote; and (b) detecting specific hybridisation of detectably
CC   labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC   having a fragment identical to the predicted exon, the probe is included
CC   in the above mentioned microarray; assigning exons to a single gene,
CC   comprising (a) identifying exons from genomic sequence by the method
CC   above and (b) measuring the expression of each of the exons in several
CC   tissues and/or cell types using hybridisation to a single exon
CC   microarrays having a probe with the exon, where a common pattern of
CC   expression of the exons in the tissues and/or cell types indicates that
CC   the exons should be assigned to a single gene; a peptide comprising one
CC   of 12011 sequences, mentioned in the specification, or encoded by the
CC   probes/open reading frames (ORF). The probes are used for gene expression
CC   analysis, and for identifying exons in a gene, particularly using human
CC   lung derived mRNA and for the study of lung diseases such as asthma, lung
CC   cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC   disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC   tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC   Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC   histiocytosis, lymphangioleiomyomtosis, pulmonary alveolar proteinosis,
CC   Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC   dyskinesis, pulmonary hypertension and hyaline membrane disease. The
CC   present sequence is a single exon probe open reading frame of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the Printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;

Query Match   94.9%; Score 1470.4; DB 6; Length 4778;
Best Local Similarity   99.9%;
Matches 1471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   78 GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG 137
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4778 GTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGG 4719

Qy   138 ATGTAAATCAGTGGTTAGAAAGTCATAAGAGTCATAAGAGTTAAAAAAAACACAGGGAAATTTGACTG 197
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4718 ATGTAAATCAGTGGTTAGAAAGTCATAAGAGTCATAAGAGTTAAAAAAAACACAGGGAAATTTGACTG 4659

Qy   198 AACAAGACGTGAATGGAGCAGTCTTGAGTGGTTAAAAAGAACATCTTGTTGATATGG 257
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4658 AACAAGACGTGAATGGAGCAGTCTTGAGTGGTTAAAAAGAACATCTTGTTGATATGG 4599

Qy   258 GCATCACACATGGACCAGCTATTCAAATGAAAGAACTATTCAAAGAATTGCGGAAAACAG 317
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4598 GCATCACACATGGACCAGCTATTCAAATGAAAGAACTATTCAAAGAATTGCGGAAAACAG 4539

Qy   318 CCATTGAAGATTCAGATCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAG 377
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4538 CCATTGAAGATTCAGATCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAG 4479

Qy   378 ACCAAACTGTGTCTCAAAAGGAACCTAGAAGAACCTAGAAGAAACTTCAAAGCAAAACAAAAAGGGTAAAG 437
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4478 ACCAAACTGTGTCTCAAAAGGAACCTAGAAGAACCTAGAAGAAACTTCAAAGCAAAACAAAAAGGGTAAAG 4419

Qy   438 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTTAAGTCAC 497
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4418 AGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTTAAGTCAC 4359

Qy   498 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGCAACCATCCATAG 557
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4358 TAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGCAACCATCCATAG 4299

Qy   558 ACCTGACATGTGTATCATATCCATTGATGAATTCAGTAATCATATCGTTACAAGTTGG 617
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4298 ACCTGACATGTGTATCATATCCATTGATGAATTCAGTAATCATATCGTTACAAGTTGG 4239

Qy   618 ATTTAGTCTACAGCCTGAAACAGGACACCAGGCAATCTCATTGATCCGATCATGAATTCA 677
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4238 ATTTAGTCTACAGCCTGAAACAGGACACCAGGCAATCTCATTGATCCGATCATGAATTCA 4179

Qy   678 AAGCCTTCACAATACAGCAACAGCCACAGAAGAGATGTCAAGATGAAATTTAGCAATG 737
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4178 AAGCCCTTCACAAATACAGCAACAGCCACAGAAGAGATGTCAAGATGAAATTTAGCAATG 4119

Qy   738 AGGTTTTCCGATTTGCTTCAGCTTGTGTATGAATTCACGTACCAATGGCACTATTCATTTTG 797
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4118 AGGTTTTCCGATTTGCTTCAGCTTGTGTATGAATTCACGTACCAATGGCACTATTCATTTTG 4059

Qy   798 GAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 857
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4058 GAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGG 3999

Qy   858 AAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTGAAGACCATCAGTCC 917
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3998 AAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTGAAGACCATCAGTCC 3939

Qy   918 AACAAGCAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTA 977
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3938 AACAAGCAAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTA 3879

Qy   978 CTCTATCTGACAGATTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT 1037
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3878 CTCTATCTGACAGATTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAAT 3819

Qy   1038 ATGATTACTTCCAGATAAAATGCAAAATTACAACACAAAATATGGGAACAAAGTAAAA 1097
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3818 ATGATTACTTCCAGATAAAATGCAAAATTACAACACAAAATATGGGAACAAAGTAAAA 3759

Qy   1098 AATTCTCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAAAATAAAGTTG 1157
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3758 AATTCTCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAAAATAAAGTTG 3699

Qy   1158 AATTCAGAGCATTTAAAGCAATTAAAGCAGATTTAAAACACTGGCAGAGTCCAGAAAGCAGCAGAAGG 1217
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3698 AATTCAGAGCATTTAAAGCAATTAAAGCAGATTTAAAACACTGGCAGAGTCCAGAAAGCAGCAGAAGG 3639

Qy   1218 AAAAATTCAGAGCAAAACAAATAAAAAAAAAGAAAACAAATAAAAAAAAAGAAAACAAAT 1277
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3638 AAAAATTCAGAGCAAAACAAATAAAAAAAAAGAAAACAAATAAAAAAAAAGAAAACAAAT 3579

Qy   1278 TGACAGGAAATCAAGATTTGTTAGATAATTCATCATGAACAGTAACAGTACCATTCTTGTAACAA 1337
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3578 TGACAGGAAATCAAGATTTGTTAGATAATTCATCATGAACAGTAACAGTACCATTCTTGTAACAA 3519

Qy   1338 ATAAATGCCACCCAGATCAAAACAAAACACTTAGATTTCCTGAAGGAAATTAAAATGGTTTG 1397
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3518 ATAAATGCCACCCAGATCAAAACAAAACACTTAGATTTCCTGAAGGAAATTAAAATGGTTTG 3459

Qy   1398 CTGTATTGGAGTTGATCCTGAGTCTCAACATCAATGAGTGGTCAAAGCTTACAAAGAAA 1457
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3458 CTGTATTGGAGTTGATCCTGAGTCTCAACATCAATGAGTGGTCAAAGCTTACAAAGAAA 3399

Qy   1458 GCCGAGTAGCAAACCTCACTTTCCACTTTCCAGTGTATATGTGAACAGAAAACCACCAAATG 1517
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3398 GCCGAGTAGCAAACCTCACTTTCCACTTTCCAGTGTATATGTGAACAGAAAACCACCAAATG 3339

Qy   1518 AGACGATTTCTACTCTAAATCTTTACCATCAA 1549
              ||||||||||||||||||||||||||||||||
Db   3338 AGACGATTTCTACTCTAAATCTTTACCATCAA 3307

RESULT 12
AAI58244
ID   AAI58244 standard; cDNA; 4771 BP.
XX

AC   AAI58244;

XX

DT   22-OCT-2001 (first entry)

XX

DE   Human polynucleotide SEQ ID NO 447.

XX

KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.

XX

OS   Homo sapiens.

XX

PN   WO200153312-A1.

XX

PD   26-JUL-2001.

XX

PF   26-DEC-2000; 2000WO-US034263.

XX

PR   23-DEC-1999;   99US-00471275.
PR   21-JAN-2000; 2000US-00488725.
PR   25-APR-2000; 2000US-00552317.
PR   20-JUN-2000; 2000US-00598042.
PR   19-JUL-2000; 2000US-00620312.
PR   03-AUG-2000; 2000US-00653450.
PR   14-SEP-2000; 2000US-00662191.
PR   19-OCT-2000; 2000US-00693036.
PR   29-NOV-2000; 2000US-00727344.

XX

PA   (HYSB-) HYSEQ INC.

XX

PI   Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI   Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang Y, Zhao QA;
PI   Zhou P, Goodrich R, Drmanac RT;

XX

DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM39088.

XX

PT   Novel nucleic acids and polypeptides, useful for treating disorders such
PT   as central nervous system injuries.

XX

PS   Claim 1; SEQ ID NO 447; 10078pp; English.

XX

CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC   encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders. Note: The sequence data for this patent did not form
CC   part of the printed specification

XX

SQ   Sequence 4771 BP; 1740 A; 833 C; 909 G; 1289 T; 0 U; 0 Other;

Query Match   94.4%;   Score 1462.4;   DB 4;   Length 4771;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches 1463;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   86 ATGGCAAAGCAACTTAACCTTCCAGAAAATACAGAAAATACAGATGATTGGACAAAAGAGGATGTAAAT 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGCAAAGCAACTTAACCTTCCAGAAAATACAGAAAATACAGATGATTGGACAAAAGAGGATGTAAAT 60

Qy   146 CAGTGGTTAGAAGTCATAAGATTGACCAAAAACACAGGGAAATTTGACTGAACAAGAC 1200
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 AGAGCAAAAACAAATAAAAAAAGAAGAGAGAGGGACCAAAGTTGGTTAAATTATTGACAGGA 1200

Db   61 CAGTGTTGTTAGAAAGTCATAGAGATTGACCAAAAACACAGGGAAATTTGACGAACAAGAC 120

Qy   206 GTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA 265
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA 180

Qy   266 CATGGACCAGCTATTCAAATAGAGAACTATTCAAAGAATTGCGGAAACAGCAGCCATTGAA 325
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CATGGACCAGCTATTCAAATAGAGAACTATTCAAAGAATTGCGGAAACAGCAGCCATTGAA 240

Qy   326 GATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGTCCTAAAGACCAAAACT 385
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGTCCTAAAGACCAAAACT 300

Qy   386 GTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAAACAAAAGGGTAAAGAGAACCCA 445
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GTGTCTTCAAAAGGAACCTAGAGAACTTCAAAGCAAAAACAAAAGGGTAAAGAGAACCCA 360

Qy   446 GATATGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTT 505
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GATATGCTAATCCGTCTGCAATGAGTACACACTGCTAAAGGTTCTAAGTCACTAAAAGTT 420

Qy   506 GAGCTCATAGAAGATAAAATAGATTATATACAAAGGAAAGGCAACCATCCATAGACCTGACA 565
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGCTCATAGAAGATAAAATAGATTATATACAAAGGAAAGGCAACCATCCATAGACCTGACA 480

Qy   566 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTTAGT 625
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTTAGT 540

Qy   626 CTACAGCCTGAAACACAGGACCAATCTCATTGATCCGATACATGATGAATTCAAAGCCTTC 685
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CTACAGCCTGAAACACAGGACCAATCTCATTGATCCGATACATGATGAATTCAAAGCCTTC 600

Qy   686 ACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAATTTAGCAATGAAGGTTTTC 745
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAATTTAGCAATGAAGGTTTTC 660

Qy   746 CGATTTGCTTGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAAA 805
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CGATTTGCTTGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAAA 720

Qy   806 GACAAACCCATGGGAAAATGTTGGCATCAAAGTCACCAATGATACCAAGGAAGCCCTC 865
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GACAAACCCATGGGAAAATGTTGGCATCAAAGTCACCAATGATACCAAGGAAGCCCTC 780

Qy   866 ATTAACCATTCAATCTGATGATAAACAAGTATTTGAAGACCATCAAGTCCAACAAGCA 925
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 ATTAACCATTTCAATCTGATGATAAACAAGTATTTGAAGACCATCAAGTCCAACAAGCA 840

Qy   926 AAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTACTCTATCT 985
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTACTCTATCT 900

Qy   986 GACAGATTGTTATTGAAGTGCACATTATTCCACAGTTCTCTGAATGCCAATATGATTAC 1045
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GACAGATTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAATATGATTAT 960

Qy   1046 TTCCAGATTAAAATGCAAAATTACAACAACAAAAATGGGAACAAGTAAAAAATTCTCA 1105
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TTCCAGATTAAAATGCAAAATTACAACAACAAAATGGGAACAAGTAAAAAATTCTCA 1020

Qy   1106 CTATTTGTGCGAGATGGGACCCAGCTCTAAGGACATTACGAAAAAATAAAGTTGATTTCAGA 1165
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CTATTTGTGCGAGATGGGACCCAGCTCTAAGGACATTACGAAAAAATAAAGTTGATTTCAGA 1080

Qy   1166 GCATTTAAAGCAGATTTAAAACACTGGCAGAGTCCAGAAAGCAGCAGAGAAAAATTC 1225
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GCATTTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAAGCAGCAGAGAAAAATTC 1140

```
RESULT 13
ADB48211
ID   ADB48211 standard; cDNA; 4771 BP.
XX
AC   ADB48211;
XX
DT   04-DEC-2003   (first entry)
XX
DE   Novel human cDNA SEQ ID NO 121.
XX
KW   ss; cancer; neurodegenerative disease; human.
XX
OS   Homo sapiens.
XX
PN   US2003104529-A1.
XX
PD   05-JUN-2003.
XX
PF   04-JAN-2002; 2002US-00037270.
XX
PR   21-JAN-2000; 2000US-00488725.
PR   25-APR-2000; 2000US-00552317.
PR   19-JUL-2000; 2000US-00620312.
XX
PA   (ZHOU/)  ZHOU P.
PA   (TANG/)  TANG Y T.
PA   (LIUC/)  LIU C.
PA   (ASUN/)  ASUNDI V.
PA   (DRMA/)  DRMANAC R T.
XX
PI   Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;
XX
DR   WPI; 2003-678194/64.
XX
PT   New polynucleotide, useful for treating diseases e.g., cancer or
PT   neurodegenerative diseases.
XX
PS   Claim 1; SEQ ID NO 121; 99pp; English.
XX
CC   The invention relates to a polynucleotide comprising a sequence given in
CC   the specification, or its mature protein-coding portion, or its
CC   complement. The polynucleotide is useful for treating diseases e.g.,
CC   cancer or neurodegenerative diseases and many others listed in the
CC   specification. The present sequence represents a novel human cDNA. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=20030104529.
XX
SQ   Sequence 4771 BP; 1740 A; 833 C; 909 G; 1289 T; 0 U; 0 Other;

Query Match   94.4%;  Score 1462.4;  DB 8;  Length 4771;
Best Local Similarity 99.9%;  Pred. No. 0;
Matches 1463;  Conservative 1;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    86 ATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAGAGGATGTAAAT 145
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAGAGGATGTAAAT 60

Qy   146 CAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAAATTTGACTGAACAAGAC 205
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAAATTTGACTGAACAAGAC 120

Qy   206 GTGAATGGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA 265
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GTGAATGGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA 180

Qy   266 CATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAGCCATTGAA 325
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAGCCATTGAA 240

Qy   326 GATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAAATGCTCCTAAAGACCAAACT 385
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAAATGCTCCTAAAGACCAAACT 300

Qy   386 GTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAAACAAAGAGGGTAAAGAGAACCCA 445
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAAACAAAGAGGGTAAAGAGAACCCA 360

Qy   446 GATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTT 505
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTT 420

Qy   506 GAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA 565
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA 480

Qy   566 TGTGTATCATATCCATTTGATGAATTCAGTTGATAATCGTTACAAGTTGGATTTTAGT 625
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TGTGTATCATATCCATTTGATGAATTCAGTTGATAATCGTTACAAGTTGGATTTTAGT 540

Qy   626 CTACAGCCTGAAACAGCAACAGGACCAGCCAATCTCATTGATCCGATACACTTCAAAGCCTTC 685
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CTACAGCCTGAAACAGCAACAGGACCAGCCAATCTCATTGATCCGATACACTTCAAAGCCTTC 600

Qy   686 ACAAATACAGCAACAGCCCAGATGTCAAGATGAAATTTAGCAATGAAATTTAGCAATGAGGTTTTC 745
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ACAAATACAGCAACAGCCCAGATGTCAAGATGAAATTTAGCAATGAAATTTAGCAATGAGGTTTTC 660

Qy   746 CGATTGCTTCGAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAAA 805
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CGATTGCTTCGAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAAA 720

Qy   806 GACAAACCCCATGGGAAAAATTGTTCGCATCAAGTCACCAATGATCCAAGGACCATCAAGCA 865
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GACAAACCCCATGGGAAAAATTGTTCGCATCAAGTCACCAATGATCCAAGGACCATCAAGCA 780

Qy   866 ATTAACCATTTCAATCGATGATAAACAACAGTATTTTGAAGACTCCAAATAGTACTCTATCT 925
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 ATTAACCATTTCAATCGATGATAAACAACAGTATTTTGAAGACTCCAAATAGTACTCTATCT 840

Qy   926 AAGAAGTGCATTCGAGAGCCAAGATTTGTGTGGAAGTTTTACTGCGCCAATAGTACTCTATCT 985
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AAGAAGTGCATTCGAGAGCCAAGATTTGTGTGGAAGTTTTACTGCGCCAATAGTACTCTATCT 900

Qy   986 GACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCCTGAATGCCAATATGATTAC 1045
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCCTGAATGCCAATATGATTAT 960

Qy  1046 TTCCAGATTAAAATGCAAAATTACAACAACAAGTAAATATGGAACAAAGTAAAAAATTCTCA 1105
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TTCCAGATTAAAATGCAAAATTACAACAACAAGTAAATATGGAACAAAGTAAAAAATTCTCA 1020

Qy  1106 CTATTTGTGCGAGATGGGACCAGCTTCAAGGACTCTCAAGGACATTACGAAAAATAAGTTGATTCAGA 1165
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  1286 AATCAAGAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTAACAAATAAATGC 1345
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 AATCAAGAGATTTGTTAGATAATTCATACTATGAACAGTACATTCTTGTAACAAATAAATGC 1260

Qy  1346 CACCCAGATCAAACAAACAAAACACTTAGATTTCCTGAAGGAAATTAAATGGTTTGCTGTATTG 1405
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CACCCAGATCAAACAAACAAAACACTTAGATTTCCTGAAGGAAATTAAATGGTTTGCTGTATTG 1320

Qy  1406 GAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAGCTTACAAAGAAAGCCGAGTA 1465
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 GAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAGCTTACAAAGAAAGCCGAGTA 1380

Qy  1466 GCAAACTTCACTTTCCAAGTGTATATGTAGAACAGAAAACCACCACCACCAATGAGACGATT 1525
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 GCAAACTTCACTTTCCAAGTGTATATGTAGAACAGAAAACCACCACCACCAATGAGACGATT 1440

Qy  1526 TCTACTCTAAATCTTTACCATCAA 1549
            ||||||||||||||||||||||||
Db  1441 TCTACTCTAAATCTTTACCATCAA 1464
```

Db 1021 CTATTTGTGCGAGATGGGACCAGCTTCTAAGGACATTACCGAAAAATAAAGTTGATTTCAGA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1166 GCATTTAAAGCAGATTTAAAACTGGCAGAGTCCAGAAAAAAGCAGCAGAAGAAAAATTC 1225
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 GCATTTAAAGCAGATTTAAAACTGGCAGAGTCCAGAAAAAAGCAGCAGAAGAAAAATTC 1140
Qy 1226 AGAGCAAAAACAAATAAAAGAAGAAAGAAGAAAGAGAGAGAGAGAGGGACCAAAGTTGGTTAAATTATTGACACGA 1285
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 AGAGCAAAAACAAATAAAAGAAGAAAGAAGAAAGAGAGAGAGAGAGGGACCAAAGTTGGTTAAATTATTGACACGGA 1200

Qy 1286 AATCAAGAGATTTGTTAGATAATTCATATCATGAACAGTACAATTCTTGTAACAAATAAATGC 1345
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 AATCAAGAGATTTGTTAGATATTTCATATCATGAACAGTACATTCTTGTAACAAATAAATGC 1260
Qy 1346 CACCCAGATCAAAACAAAACACTTAGATTTCCTGAAGGAATTAAATGGTTGCTGTATTG 1405
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 CACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAATTAAATGGTTTGCTGTATTG 1320
Qy 1406 GAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAAAGCCGAGTA 1465
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 GAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAAAGCCGAGTA 1380
Qy 1466 GCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAACCACCACCAAATGAGACGACGATT 1525
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 GCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAACCACCACCAAATGAGACGACGATT 1440
Qy 1526 TCTACTCTAAATCTTTACCATCAA 1549
          |||||||||||||||||||||||
Db 1441 TCTACTCTAAATCTTTACCATCAA 1464

RESULT 14
AAS01147
ID AAS01147 standard; DNA; 6233 BP.
XX
AC AAS01147;
XX
DT 12-SEP-2001 (first entry)
XX
DE Interferon induced nucleic acid, IFN2.
XX
KW Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
KW immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
KW graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
KW human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
KW haematologic disease; chronic neutropenia; myocardial infarction;
KW neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
KW amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN2; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   1..4770
FT   /tag= a
FT   /product= "IFN2"
FT   /note="Interferon induced polypeptide"
XX
PN WO20011B208-A2.
XX
PD 15-MAR-2001.
XX
PF 08-SEP-2000; 2000WO-US024704.
XX
PR 08-SEP-1999;   99US-0152921P.
PR 20-OCT-1999;   99US-0160575P.
PR 20-JAN-2000;   2000US-0177104P.
PR 07-SEP-2000;   2000US-00656633.
XX
PA (CURA-) CURAGEN CORP.
PA (BIOJ) BIOGEN INC.
XX
PI Peyman JA, Da Silva A, Hochman P, Hsu A;
XX

DR WPI; 2001-235201/24.
DR P-PSDB; AAU00294.
XX
PT New interferon induced polypeptides and polynucleotides, useful for the
PT diagnosis, prevention and treatment of immunological, cell proliferative
PT disorders, such as lupus erythematosus, cancer, stroke and Alzheimer's
PT disease.
XX
PS Claim 9; Page 17-22; 134pp; English.
XX
CC The sequence represents interferon induced nucleic acid, IFN2. IFN
CC nucleic acids and polypeptides are useful for treating or preventing a
CC pathology associated with IFN polypeptide in a human. They are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
CC acids, polypeptides and antibodies are useful for diagnosis, prevention
CC or treatment of variety of immunological and cell proliferative
CC disorders, such as autoimmune diseases-e.g. lupus erythematosus,
CC immunodeficiency diseases such as acquired immunodeficiency syndrome
CC (AIDS), graft rejection, viral infections including hepatitis and human
CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
CC haematologic diseases such as aplastic anaemia and chronic neutropenia
CC and cancer. In addition they are also useful for treating or diagnosing
CC various disorders associated with cell death, including myocardial
CC infarction, stroke, neurological diseases including Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
CC atrophy. IFN nucleic acids and polypeptides are also useful for
CC identifying interferon-like proteins and interferon agonists, for
CC screening drugs and compounds which inhibit or enhance IFN activity or
CC function and as targets for the identification of small molecules that
CC are immunostimulatory, immunosuppressive or stimulate or suppress normal
CC cell or tumour cell growth in mammals, including humans
XX
SQ Sequence 6233 BP; 2290 A; 1095 C; 1157 G; 1691 T; 0 U; 0 Other;

Query Match   94.4%; Score 1462.4; DB 4; Length 6233;
Best Local Similarity   99.9%; Pred. No. 0;
Matches 1463; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

Qy   86 ATGGCAAAGCAACTTAACCTTCCAGAAATCACAGATGATTGACAAAGAGGATGTAAAT 145
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGCAAAGCAACTTAACCTTCCAGAAATCACAGATGATTGACAAAGAGGATGTAAAT 60
Qy 146 CAGTGGTTAGAAAGTCATAAGAGTTGACCAAAACACAGGGAAATTTGACTGAACAAGAC 205
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CAGTGGTTAGAAAGTCATAAGAGTTGACCAAAACACAGGGAAATTTGACTGAACAAGAC 120
Qy 206 GTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA 265
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 GTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA 180
Qy 266 CATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAACAGCCATTGAA 325
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAACAGCCATTGAA 240
Qy 326 GATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAGACCAAACT 385
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAAAGACCAAACT 300
Qy 386 GTGTCTCAAAAGGAACGTAGAGAACGTAGAGAAACTTCAAAGACAAAACAAAAGGGTAAAGAACCCA 445
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GTGTCTCAAAAGGAACGTAGAGAACGTAGAGAAACTTCAAAGACAAAACAAAAGGGTAAAGAACCCA 360
Qy 446 GATATGGCTAATCCGTCGCAATGAGTACAACTGCTAAGGTTCTAAGGTTCTACTCAAAGTT 505
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 GATATGGCTAATCCGTCTGCAATGAGTGTCAATGAGTACAACTGCTAAGGTTCTACTCAAAGTT 420
Qy 506 GAGCTCATAGAAGATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA 565
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 GAGCTCATAGAAGATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA 480
Qy 566 TGTGTATCCATATCCATTTGATGAATTCAGTAGTGTTACAAGTTGGATTTTAGT 625
         |||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db       481  TGTCTATCATATCCATTCATTTGATGAATTCAGTAATCCATATCGTTACACAGTTGGATTTTAGT  540

Qy       626  CTACAGCCTGAAACAGAGACCAGGCAATCTCATTGATCCGATACATGAATTCAAAGCCTTC  685
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541  CTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCAAAGCCTTC  600

Qy       686  ACAAATACAGCAACAGCACCAGAAGAGAGATGTCAAGATGAAATTTAGGCAATGAGGTTTTC  745
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601  ACAAATACAGCAACAGCACCAGAAGAGAGATGTCAAGATGAAATTTAGCAATGAGGTTTTTC  660

Qy       746  CGATTTGCTTCAGCTTGTGATGAATTCACGTACCAATGGCACTATTCATTTGGAGTCAAA  805
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661  CGATTTGCTTCAGCTTGTGATGAATTCACGTACCAATGGCACTATTCATTTGGAGTCAAA  720

Qy       806  GACAAACCCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGGAAGCCCTC  865
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721  GACAAACCCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGGAAGCCCTC  780

Qy       866  ATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCAACAAGCA  925
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781  ATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCAACAAGCA  840

Qy       926  AAGAAGTGCATTGGAGAGCCAAGATTTGTGGGAAGTTTACTGCCAAATAGTACTCCATCT  985
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841  AAGAAGTGCATTCGAGAGCCAAGATTTGTGGGAAGTTTTACTGCCAAATAGTACTCCTATCT  900

Qy       986  GACAGATTTGTTATTGAAGTGGCACATTATTCCACAGTTCTCTGAATGCCAATATGATTAC  1045
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901  GACAGATTTGTTATTGAAGTGGCACATTATTCCACAGTTCTCTGAATGCCAATATGATTAT  960

Qy      1046  TTCCAGATTAAAATGCAAAATCAACACAAATATGGGAACAAAGTAAAAAATTCTCA  1105
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961  TTCCAGATTAAAATGCAAAATCAACACAAATATGGGAACAAAGTAAAAAATTCTCA  1020

Qy      1106  CTATTTGTGCGAGATGGGACCAGCTCTAAGGACGTCTCAAGGAGCATTAGGAAAGTTGATTTCAGA  1165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021  CTATTTGTGCGAGATGGGACCAGCTCTAAGGACGTCTCAAGGAGCATTAGGAAAGTTGATTTCAGA  1080

Qy      1166  GCATTTAAAGCAGATTTTAAAAACACTGGCAGAGTCCAGAAAAGCAGCAGAAAAATTC  1225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081  GCATTTAAAGCAGATTTTAAAAACACTGGCAGAGTCCAGAAAAGCAGCAGAAAAATTC  1140

Qy      1226  AGAGCAAAAAACAAATAAAAGAAAGAGAGGGGACCCAAGTTGGTTAAATTATTGACAGGA  1285
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141  AGAGCAAAAAACAAATAAAAGAAAGAGAGGGGACCCAAGTTGGTTAAATTATTGACACGGA  1200

Qy      1286  AATCAAGATTTGTTAGATAATTCATACTATGAAGCTTTACAAAGAAATGC  1345
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201  AATCAAGATTTGTTAGATAATTCATACTATGAAGCTTTACAAAGAAATGC  1260

Qy      1346  CACCCAGATCAAACAAAACACTTAGAGTTCCTGAAGGAATTAAAATGGTTTGCTGTATTG  1405
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261  CACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAATTAAAATGGTTTGCTGTATTG  1320

Qy      1406  GAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAAAGCCGAGTA  1465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321  GAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAAAGCCGAGTA  1380

Qy      1466  GCAAACCTTCCACTTTCCAAGTGTATATGTGTAGAGAAACCACCAAATGAGACGATT  1525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381  GCAAACCTTCCACTTTCCAAGTGTATATGTGTAGAGAACAGAAATGAGACGATT  1440

Qy      1526  TCTACTCTAAATCTTACCATCAA  1549
              |||||||||||||||||||||||
Db      1441  TCTACTCTAAATCTTACCATCAA  1464
```

RESULT 15
AAK77710
ID   AAK77710 standard; DNA; 1045 BP.
XX
AC   AAK77710;
XX

```
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32522.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0254447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
```

PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0234984P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0235327P.
PR   29-SEP-2000;   2000US-0235367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240060P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251998P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251866P.

PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 32522; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 1045 BP; 410 A; 187 C; 196 G; 252 T; 0 U; 0 Other;

Query Match   67.4%; Score 1043.4; DB 4; Length 1045;
Best Local Similarity   99.9%;
Matches 1044;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   206 GTGAATGGAGCAGTCTTGAAGTGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTGAATGGAGCAGTCTTGAAGTGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA 60

Qy   266 CATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAGCCATTGAA 325
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAGCCATTGAA 120

Qy   326 GATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAATGTCTCCTAAAGACCAAACT 385
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAATGTCTCCTAAAGACCAAACT 180

Qy   386 GTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACAAAAGGGTAAAGAAGAACCCA 445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACAAAAGGGTAAAGAAGAACCCA 240

Qy   446 GATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTT 505
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTT 300

Qy   506 GAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA 565
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA 360

Qy   566 TGTGTATCATATCCATTGATGAATTCAGTAGTTAATCCATATCGTTACAAGTTGGATTTTAGT 625
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGTGTATCATATCCATTGATGAATTCAGTAGTTAATCCATATCGTTACAAGTTGGATTTTAGT 420

Qy   626 CTACAGCCTGAAACAGGACCAGCCAATCTCATTGATCCGGATACAATGAATCAAAGCCTTC 685
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CTACAGCCTGAAACAGGACCAGCCAATCTCATTGATCCGGATACAATGAATCAAAGCCTTC 480

```
Qy   686 ACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAATTTAGCAATGAGGTTTTC 745
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAATTTAGCAATGAGGTTTTC 540

Qy   746 CGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAAA 805
         |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   541 CGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAAA 600

Qy   806 GACAAACCCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGGAAGCCCTC 865
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GACAAACCCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCAAGGAAGCCCTC 660

Qy   866 ATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCAACAAGCA 925
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCAACAAGCA 720

Qy   926 AAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTACTCTATCT 985
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 AAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTACTCTATCT 780

Qy   986 GACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAATATGATTAC 1045
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   781 GACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAATATGATTAT 840

Qy  1046 TTCCAGATTAAAATGCAAATTACAACAACAAATATGGGAACAAAGTAAAAAAATTCTCA 1105
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 TTCCAGATTAAAATGCAAATTACAACAACAAATATGGGAACAAAGTAAAAAAATTCTCA 900

Qy  1106 CTATTTGTGCGAGATGGACCAGCTCTAAGGACATTACGAAAAAATAAAGTTGATTTCAGA 1165
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 CTATTTGTGCGAGATGGACCAGCTCTAAGGACATTACGAAAAAATAAAGTTGATTTCAGA 960

Qy  1166 GCATTTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAAAGCAGCAGAGAAAAATTC 1225
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GCATTTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAAAGCAGCAGAGAAAAATTC 1020

Qy  1226 AGAGCAAAAACAAATAAAAAGAAA 1250
         ||||||||||||||||||||||||
Db  1021 AGAGCAAAAACAAATAAAAAGAAA 1045
```

Search completed: April 25, 2004, 08:46:12
Job time : 417.576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 23:54:27 ; Search time 4078.84 Seconds
(without alignments)
16460.143 Million cell updates/sec

Title: US-10-051-835-17
Perfect score: 1549
Sequence: 1 aacacatcagatatttcag.........ctctaaatctttaccatcaa 1549

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1533.8	99.0	3956	9	AK122951	AK122951 Homo sapi
2		1472.4	95.1	6826	9	HSM800722	BX647072 Homo sapi
3		1472.4	95.1	6827	9	AB095925	AB095925 Homo sapi
4	c	1472	95.0	191191	9	AC000119	AC000119 Homo sapi
5	c	1464	94.5	172371	9	AC092871	AC092871 Pan trogl
6		1462.4	94.4	4771	6	AR338630	AR338630 Sequence
7		1462.4	94.4	6233	2	AX098228	AX098228 Sequence
8	c	1396.8	90.2	176350	2	AC092513	AC092513 Papio anu
9	c	1048.6	67.7	175424	2	AC105435	AC105435 Sus scrof
10	c	1039.8	67.1	108538	2	AC108893	AC108893 Bos tauru
11	c	921	59.5	148278	2	AC079885	AC079885 Rattus no
12	c	921	59.5	180665	2	AC084240	AC084240 Rattus no
13	c	921	59.5	221557	2	AC134055	AC134055 Rattus no
14	c	788.4	50.9	119266	2	AC108900	AC108900 Felis cat
15		759	49.0	225885	2	AC113236	AC113236 Canis fam
16		672	43.4	705	6	AX385208	AX385208 Sequence
17		636.8	41.1	7122	9	AB095926	AB095926 Homo sapi
18		628.4	40.6	5870	6	AX098226	AX098226 Sequence
19	c	599.6	38.7	137638	2	AC110666	AC110666 Canis fam
20	c	499	32.2	188579	2	AC107316	AC107316 Mus muscu
21	c	498	32.1	195282	2	AC022453	AC022453 Mus muscu
22	c	473.8	30.6	125962	2	AC084271	AC084271 Rattus no
23	c	473.8	30.6	191284	2	AC129122	AC129122 Rattus no
24		469	30.3	262697	8	AC108642	AC108642 Rattus no
25	c	314.6	20.3	52652	2	AC104854	AC104854 Homo sapi
26		169.2	10.9	188729	2	BX005331	BX005331 Danio rer
27		138	8.9	356	2	AX245250	AX245250 Sequence
28	c	97.8	6.3	262697	6	AC108642	AC108642 Rattus no
29	c	95.2	6.1	147824	5	BX548026	BX548026 Danio rer
30		95.2	6.1	157603	5	AL953865	AL953865 Zebrafish
31		73	4.7	251	2	AX246569	AX246569 Sequence
32	c	72.2	4.7	1453	8	AJ591978	AJ591978 Arabidops
33	c	69.6	4.5	1434	5	AJ592058	AJ592058 Arabidops
34	c	69.4	4.5	139627	5	AL929509	AL929509 Zebrafish
35		68.6	4.4	7218	6	I66494	I66494 Sequence 14
36	c	68.2	4.4	93791	2	AC138073	AC138073 Homo sapi
37		67.8	4.4	188355	2	BX537350	BX537350 Danio rer
38		66.4	4.3	347050	2	PFA929351	AL929351 Plasmodiu
39	c	66.2	4.3	60604	2	AC023466	AC023466 Homo sapi
40		66	4.3	1407	8	AJ592026	AJ592026 Arabidops
41	c	63	4.1	188477	2	BX470139	BX470139 Danio rer
42	c	61.6	4.0	349980	6	AX344567	AX344567 Sequence
43	c	61	3.9	14006	2	AX346860	AX346860 Sequence
44	c	60.8	3.9	128683	2	AC135793	AC135793 Oryza sat
45	c	60.8	3.9	143585	2	AC013349	AC013349 Homo sapi

ALIGNMENTS

RESULT 1
AK122951
LOCUS   AK122951   3956 bp   mRNA   linear   PRI 09-SEP-2003
DEFINITION Homo sapiens cDNA FLJ16668 fis, clone THYMU2039788.
ACCESSION   AK122951
VERSION   AK122951.1 GI:34528396
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
            Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J.,

Pred. No. is the number of results predicted by chance to have a

```
           Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K.,
           Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H.,
           Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A.,
           Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
           and Isogai,T.
  TITLE    NEDO human cDNA sequencing project
  JOURNAL  Unpublished
REFERENCE  2  (bases 1 to 3956)
  AUTHORS  Isogai,T. and Yamamoto,J.
  TITLE    Direct Submission
  JOURNAL  Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
           Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
           (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
  COMMENT  NEDO human cDNA sequencing project supported by Ministry of
           Economy, Trade and Industry of Japan; cDNA full insert sequencing:
           Research Association for Biotechnology (RAB); cDNA library
           construction: Helix Research Institute (HRI) (supported by Japan
           Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
           HRI, and Biotechnology Center, National Institute of Technology and
           Evaluation: clone selection for full insert sequencing: HRI and
           RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..3956
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="THYMU2039788"
                     /tissue_type="thymus"
                     /clone_lib="THYMU2"
                     /note="cloning vector: pME18SFL3"
ORIGIN

Query Match     99.0%;  Score 1533.8;  DB 9;  Length 3956;
Best Local Similarity  99.8%;  Pred. No. 1.2e-279;
Matches 1546;  Conservative   0;  Mismatches  2;   Indels  1;   Gaps   1;

Qy       1 AACACATCAGATATTTCAGCACTAAAGAGATGGTTTCCCCACATATATGTAAAAGAA  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24 AACACATCAGATATTTCAGCACTAAAGAGATGGTTTCCCCACATATATGTAAAAGAA  83

Qy      61 ATTTGCAAGACTACTGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGA  120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84 ATTTGCAAGACTACTGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGA  143

Qy     121 TGATTGGACAAAAGGATGGATGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAACA  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     144 TGATTGGACAAAAGGATGGATGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAACA  203

Qy     181 CAGGGAAATTTGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGTTAAAAAAGA  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     204 CAGGGAAATTTTGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGTTAAAAAAGA  263

Qy     241 ACATCTTGTTGATATGGGCATCACACATGGACCACAGCTATTCAAATAGAAGAACTATTCAA  300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     264 ACATCTTGTGATATGGGCATCACACATGGACCACAGCTATTCAAATAGAAGAACTATTCAA  323

Qy     301 AGAATTGCGGAAAACAGCATTGAAGATTCGATTCAGACATCTAAGATGGGAAGCCCAG  360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     324 AGAATTGCGGAAAACAGCATTGAAGATTCGATTCAGACATCTAAGATGGGAAGCCCAG  383

Qy     361 TAAAAATGTCCTAAAGACCAAACTTGTCTCAAAAGAAGTAGAGAAACTTCAAAGCA  420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     384 TAAAAATGTCCTAAAGACCAAACTTGTCTCAAAAGAAGTAGAGAAACTTCAAAGCA  443

Qy     421 AAACAAAAGGGTAAAGAAAGAGAACCCAGATAATGGCTAATCCGTCTGCAATGACTGC  480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     444 AAACAAAAGGGTAAAGAAAGAGAACCCAGATAATGGCTAATCCGTCTGCAATGAGTACAACTGC  503

Qy     481 TAAAGGTTCTAGTCACTAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGA  540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     504 TAAAGGTTCTAGTCACTAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGA  563

Qy     541 AAGGCAACCATCCATAGACCTGACATGTGTATCATCATTTGATGAATTCAGTAATCC  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     564 AAGGCAACCATCCATAGACCTGACATGTGTATCATCATTTGATGAATTCAGTAATCC  623

Qy     601 ATATCGTTACAAGTTGGATTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGA  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     624 ATATCGTTACAAGTTGGATTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGA  683

Qy     661 TCCGATACATGAATTCAAAGCTTCACAAATACAGCAACAGCCAACAGAGGATGTCAA  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     684 T-CGATACATGAATTCAAAGCTTCACAAATACAGCAACAGCCAACAGAGGATGTCAA  742

Qy     721 GATGAAATTTAGCAATGAGGTTTCGATTGCTTCAGCTTGTATGAATTCACGTACCAA  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     743 GATGAAATTTAGCAATGAGGTTTCGATTGCTTCAGCTTGTATGAATTCACGTACCAA  802

Qy     781 TGGCACTATTCATTTGGAGTCAAAGACAAACCCATGGCAAAATTGTTGGCATCAAAGT  840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     803 TGGCACTATTCATTTGGAGTCAAAGACAAACCCCATGGGCAAAATTGTTGGCATCAAAGT  862

Qy     841 CACCAATGATACCAAGGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTT  900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     863 CACCAATGATACCAAGGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTT  922

Qy     901 TGAAGACCATCAAGTCCAACAGAAGCAAAGAAGTGCATTCGAGAGCCAAGAGTTGTGAAGT  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     923 TGAAGACCATCAAGTCCAACAGAAGCAAAGAAGTGCATTCGAGAGCCAAGAGTTGTGAAGT  982

Qy     961 TTTACTGCCAAATAGTACTCTCATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACA  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     983 TTTACTGCCAAATAGTACTCTCATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACA  1042

Qy    1021 GTTCTCTGAATGCCAATATGATTACTTCCAGATTAATGATGATTATTTCCAGATTAATGCCAATGTCAAAATTACACACAAAAT  1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1043 GTTCTCTGAATGCCAATATGATTATTTCCAGATTAATGCCAATGTCAAAATTACACACAAAAT  1102

Qy    1081 ATGGGAACAAAGTAAAAAATTCTCACTATTTGTGCGAGATGGACCCAGCTCTAAGGACAT  1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1103 ATGGGAACAAAGTAAAAAATTCTCACTATTTGTGCGAGATGGACCCAGCTCTAAGGACAT  1162

Qy    1141 TACGAAAATAAAGTTGATTTCAGAGCATTTAAAGCAGATTTTAAAACACTGGCAGAGTC  1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1163 TACGAAAATAAAGTTGATTTCAGAGCATTTAAAGCAGATTTTAAAACACTGGCAGAGTC  1222

Qy    1201 CAGAAAAGCACGCAGGAAGAAATTCAGAGCAAGAAATTTCAGGACAGAAACAAATAAAAAGAGAGGGACC  1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1223 CAGAAAAGCACGCAGGAAGAAATTTCAGGACAGAAACAAATAAAAAGAGAGGGACC  1282

Qy    1261 AAAGTTGGTTAAATTATTGACAGGAATCAAGATTTGTTAGATAATTCATACTATGAACA  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1283 AAAGTTGGTTAAATTATTGACAGGAATCAAGATTTGTTAGATAATTCATACTATGAACA  1342

Qy    1321 GTACATTCTTGTAACAAATAAATGCCACCCAGTCAAACAAACAAAACACTTAGATTTCCTGAA  1380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1343 GTACATTCTTGTAACAAATAAATGCCACCCAGTCAAACAAACAAAACACTTAGATTTCCTGAA  1402

Qy    1381 GGAAAATAAATGTTTGCTGTATTGGAGTTGATCCTGAGTCTAACATCAATCAATGGAGTGGT  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1403 GGAAAATAAATGTTTGCTGTATTGGAGTTGATCCTGAGTCTAACATCAATCAATGGAGTGGT  1462

Qy    1441 CAAAGCTTACAAGAAGCCGAGTAGCAAACTTCACTTTCCAAGTGTATATGTAGAACA  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1463 CAAAGCTTACAAGAAGCCGAGTAGCAAACTTCACTTTCCAAGTGTATATGTAGAACA  1522

Qy    1501 GAAAACCACCACCAAATGAGACGATTCTCACTCTAAATCTTTACCATCAA  1549
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1523 GAAAACCACCACCAAATGAGACGATTCTCACTCTAAATCTTTACCATCAA  1571


RESULT 2
HSM808722
LOCUS       HSM808722              6826 bp    mRNA    linear   PRI 30-AUG-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686B0866 (from clone DKFZp686B0866).
ACCESSION   BX647072
```

VERSION   BX647072.1 GI:34367552
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 6826)
AUTHORS   Bkecker,H., Boecher,M., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
CONSRTM   The German Human cDNA Consortium
TITLE   Direct Submission
JOURNAL   Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp686B0866) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES   Location/Qualifiers
     source   1..6826
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /map="7q21-7q22"
                     /clone="DKFZp686B0866"
                     /tissue_type="human endometrium carcinoma cell line"
                     /clone_lib="686 (synonym: hlcc3). Vector pSporti_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
     polyA_signal   6789..6794
     polyA_site   6802
ORIGIN

Query Match   95.1%;   Score 1472.4; DB 9; Length 6826;
Best Local Similarity   99.6%;   Pred. No. 4.2e-268;
Matches 1476;   Conservative   0;   Mismatches   6;   Indels   0;   Gaps   0;

Qy   68 AGACTACTGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGAAGATGATTGG 127
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   209 AGTCAGTTTGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGAAGATGATTGG 268

Qy   128 ACAAAAGAGGATGTAAATCAGTGTTAGAAAGTCATAAAGTCATAAAGAATGACCAAAACACAGGGAA 187
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   269 ACAAAAGAGGATGTAAATCAGTGTTAGAAAGTCATAAAGTCATAAAGAATGACCAAAACACAGGGAA 328

Qy   188 ATTTGACTGAACAAGACGTGAATGGACCAGTGCAGTCTTGAAGTGGTTAAAAAAGAACATCTT 247
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   329 ATTTGACTGAACAAGACGTGAATGGACCAGTGCAGTCTTGAAGTGGTTAAAAAAGAACATCTT 388

Qy   248 GTTGATATGGGCATCACACAGAACCCAGATATGGCTAATCCGTCTGAAGATGGGAAAGCCAGTAAAGAATTG 307
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   389 GTTGATATGGGCATCACACAGAACCCAGATATGGCTAATCCGTCTGAAGATGGGAAAGCCAGTAAAGAATTG 448

Qy   308 CGGAAAACAGCCATTGAAGATTCAGATCAGATCAGATCTAAGAGTCAGTAGAGAAAGGAACAACTTCAAAAAAT 367
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   449 CGGAAAAACAGCCATTGAAGATTCAGATCAGATCAGATCTAAGAGTCAGTAGAGAAAGCCCAGTAAAAAAT 508

Qy   368 GCTCCTAAAGACCAAACTGTGTCTCATGGAAGACAGTCTCAAAAGGAACTTCAAAGCAAAAACAA 427
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   509 GCTCCTAAAGACCAAACTGTGTCTCATGGAAGACAGTCTCAAAAGGAACTTCAAAGCAAAAACAA 568

Qy   428 AAGGGTAAAGAAGAAGACCCAGATATGGCTAATCCGTCTAATCCGTCTGAAGATAAAATAGATTATACAAAGGAAAGGCAA 487
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   569 AAGGGTAAAGAAGAGAACCCAGATATGGCTAATCCGTCTAATCCGTCTGAAGATATAGATTATACAAAGGAAAGGCAA 628

Qy   488 TCTAAGTCACTAAAAGTTGAGCTCATGAAGATAAAATAGATTATACAAAGGAAAGGCAA 547
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   629 TCTAAGTCACTAAAAGTTGAGCTCACAGAGAAGATAAAATAGATTATACAAAGGAAAGGCAA 688

Qy   548 CCATCCATAGACCTGACATGTGTATCATATCCATTGTGATGAATTCAGTAATCCATATCGT 607
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   689 CCATCCATAGACCTGACATGTGTATCATATCCATTGTGATGAATTCAGTAATCCATATCGT 748

Qy   608 TACAAGTTGGATTTAGTCTACAGGCTGAAACACGACCAGGCAATCTCATTGATCCGATA 667
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   749 TACAAGTTGGATTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATA 808

Qy   668 CATGAATTCAAAGCCTTCAAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAA 727
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   809 CATGAATTCAAAGCCTTCAAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAA 868

Qy   728 TTTAGCAATGAGGTTTCCGATTTGCTTCAGCTGTATGAATTCACGTACCAATGGCACT 787
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   869 TTTAGCAATGAGGTTTCCGATTTGCTTCAGCTGTATGAATTCACGTACCAATGGCACT 928

Qy   788 ATTCATTTGGAGTCAAAGACAAACCCCATGGGAAAATTGTTGGGATCAAAGTCACCAAT 847
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   929 ATTCATTTGGAGTCAAAGACAAACCCCATGGGAAAATTGTTGGGCATCAAAGTCACCAAT 988

Qy   848 GATACCAAGGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGAC 907
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   989 GATACCAAGGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGAC 1048

Qy   908 CATCAAGTCCAACAGCAAAGAAGTGCATTCGAGAGCGCAAAGAGTTGTGGGAAGTTTTACTG 967
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1049 CATCAAGTCCAACAGCAAAGAAGTGCATTCGAGAGCGCAAAGATTTGTGGGAAGTTTTACTG 1108

Qy   968 CCAAATAGTACTCTATCTGACAGATTTGTATTGGAAGTGGACATTATTCCACAGTTCTCT 1027
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1109 CCAAATAGTACTCTATCTGACAGATTTGTATTGGAAGTGGACATTATTCCACAGTTCTCT 1168

Qy   1028 GAATGCCAATATGATTACTTCCAGATTAAAATGCAAAATTACAACAACAAATATGGGAA 1087
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1169 GAATGCCAATATGATTACTTCCAGATTAAAATGCAAAATTACAACAACAAATATGGGAA 1228

Qy   1088 CAAAGTAAAAAATTCTCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAA 1147
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1229 CAAAGTAAAAAATTCTCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAA 1288

Qy   1148 AATAAAGTTGATTTGCAGAGCATTTAAAGACAGATTTTAAAACACTGGCAGAGTCCAGAAAA 1207
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1289 AATAAAGTTGATTTGCAGAGCATTTAAAGACAGATTTTAAAACACTGGCAGAGTCCAGAAAA 1348

Qy   1208 GCAGCGAAGAAAAATTCAGACGAAGAAACAAATAAAGAGAGGACCAAAGTTG 1267
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1349 GCAGCGAAGAAAAATTCAGACGAAGAAACAAATAAAGAGAGGACCAAAGTTG 1408

Qy   1268 GTTAAATTATTGACAGGAAATCAAGATTTGTTAGATAATTCATCATGAACAGTACATT 1327
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1409 GTTAAATTATTGACAGGAAATCAAGATTTGTTAGATAATTCATCATGAACAGTACATT 1468

Qy   1328 CTTGTAACAAATAAATGCCACCCAGATCAAACAAACAAAACACTTAGATTTCCTGAAGGAAATT 1387
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1469 CTTGTAACAAATAAATGCCACCCAGATCAAACAAACAAAACACTTAGATTTCCTGAAGGAAATT 1528

Qy   1388 AAATGGTTGCTGTATTGGAGTTTGATCAATCAATGGGTCAAAGCT 1447
            |||||||||||||||||||||||||||||||||||||||||||||
Db   1529 AAATGGTTGCTGTATTGGAGTTTGATCAATCAATGGGTCAAAGCT 1588

Qy   1448 TACAAAGAAAGCCGAGTAGAGAAACCTTCACTTTCCAAGTGTATATGTGAGAACAGAAAACC 1507
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1589 TACAAAGAAAGCCGAGTAGAGAAACCTTCACTTTCCAAGTGTATATGTGAGAACAGAAAACC 1648

Qy   1508 ACACCAAATGAGACGATTTCTACTCAAATCTTTACCATCAA 1549
            |||||||||||||||||||||||||||||||||||||||||
Db   1649 ACACCAAATGAGACGATTTCTACTCAAATCTTTACCATCAA 1690

RESULT 3
AB095925
LOCUS   AB095925   6827 bp   mRNA   linear   PRI 12-NOV-2002

Db	231	AGTCAGTTTGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGG 290
Qy	128	ACAAAAGAGGATGTAAATCAGTGGTTAGAAGTCATAGAAGTGATTGACCAAAAACACAGGAA 187
Db	291	ACAAAAGAGGATGTAAATCAGTGGTTAGAAGTCATAGAAGTGATTGACCAAAAACACAGGGAA 350
Qy	188	ATTTTGACTGAACAAGACGTGAATGGAGCAGTGTCTTGAAGTGTTAAAAAAAGAAACATCTT 247
Db	351	ATTTTGACTGAACAAGACGTGAATGGAGCAGTGTCTTGAAGTGGTTAAAAAAAGAAACATCTT 410
Qy	248	GTTGATATGGGCATCACACATGGACCAGCAGTCTATTCAAATAGAGAACTATTCAAAGAATTG 307
Db	411	GTTGATATGGGCATCACACATGGACCAGCAGTCTATTCAGACATTCAAATAGAAGAACTATTCAAAGAATTG 470
Qy	308	CGGAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAAT 367
Db	471	CGGAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAAT 530
Qy	368	GCTCCTAAAGACCAAACTGTGTCTCAAAAGAACGTAGAGAAACTTCAAAGCAAAAACAA 427
Db	531	GCTCCTAAAGACCAAACTGTGTCTCAAAAGAACGTAGAGAAACTTCAAAGCAAAAACAA 590
Qy	428	AAGGGTAAAGAGAACCCCAGATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGT 487
Db	591	AAGGGTAAAGAGAACCCCAGATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGT 650
Qy	488	TCTAAGTCACTAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAAGGCAA 547
Db	651	TCTAAGTCACTAAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAAGGCAA 710
Qy	548	CCATCCATAGACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGT 607
Db	711	CCATCCATAGACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGT 770
Qy	608	TACAAGTTGGATTTAGTCTACAGCCTGAAACAGGAGGACCAGGCAATCCATTGATCCGATA 667
Db	771	TACAAGTTGGATTTAGTCTACAGCCTGAAACAGGAGGACCAGGCAATCCATTGATCCGATA 830
Qy	668	CATGAATTCAAAGCCTTCACAAATACAGCAACAGCCACAGAGAGAGGATGTCAAGATGAAA 727
Db	831	CATGAATTCAAAGCCTTCACAAATACAGCAACAGCCACAGAGAGAGGATGTCAAGATGAAA 890
Qy	728	TTTAGCAATGAGGTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACT 787
Db	891	TTTAGCAATGAGGTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACT 950
Qy	788	ATTCATTTTGGAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAAT 847
Db	951	ATTCATTTTGGAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAAT 1010
Qy	848	GATACCAAGGAGGAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAGAC 907
Db	1011	GATACCAAGGAGGAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAGAC 1070
Qy	908	CATCAAGTCCAACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTTGGAAGTTTTACTG 967
Db	1071	CATCAAGTCCAACAAGCAAAGAAGTGCATTCGAGAGCCAAGATTTGTTGGAAGTTTTACTG 1130
Qy	968	CCAAATAGTACTTCATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCT 1027
Db	1131	CCAAATAGTACTTCATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCT 1190
Qy	1028	GAATGCCAATATGATTACTTCAGATTAAAATGCAAAATGCAACAACAAAATATGGGAA 1087
Db	1191	GAATGCCAATATGATTACTTCAGATTAAAATGCAAAATGCAACAACAAAATATGGGAA 1250
Qy	1088	CAAAGTAAAAAAATTCTCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAA 1147
Db	1251	CAAAGTAAAAAAATTCTCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAA 1310
Qy	1148	AATAAAGTTGATTTCAGAGCAGTTTAAAACACTGGCAGTCCAGAAAA 1207
Db	1311	AATAAAGTTGATTTCAGAGCAGTTTAAAACACTGGCAGTCCAGAAAA 1370

FEATURES   Location/Qualifiers
     source   1..6827
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="eg01144"
                     /cell_type="normal human aortic endothelial cells"
                     /note="vector:pBC SK plus"
     gene   1..6827
                     /gene="KIAA2004"
     CDS   <231..5018
                     /gene="KIAA2004"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="KIAA2004 protein"
                     /protein_id="BAC23101.1"
                     /db_xref="GI:24899174"
                     /translation="SQFGIRMAKQLNLPENTDDWTKEDVNQWLESHKIDQKHREILTE
                     QDVNGAVLKWLKKEHLVDMGITHGPAIQIEELFKELRKTAIEDSIQTSKMGKPSKNAP
                     KDQTVSKQKERRETSKQKQKGKENPDMANPSAMSTTAKGSKSKSLKVELIEDKIDYTKRQ
                     PSIDLTCVSYPPDEFSNPYRYKLDFSLQPETGPGNLIDPIHEFKAFTNTATATEEDVK
                     MKFSNEVFRFASACMNSRINGTIHFGVGKDKPHGKIVGLKVTNDTKEALIYHFNLMINK
                     YFEDHQVQQAKCIREPRFVEVLLPNSTLSDRFVIEVDIIPQFSECQYDYFQIKMQNY
                     NNKIWEQSKKFSLFVRDGTSSKDITKNKVDFRAFKADFKTLAESRKAAEEKFRAKTNK
                     KEREGPKLVKLLTGNQDLLDNSYEQYILVTNKCHPDQTKHLDFLEKIKWFAVLEFDP
                     ESNINGVVKAYKESRVANLHFPSVYVEQKTTPNETISTLNLYHQPSWIFCNGRLDLDS
                     EKYKPFDPSSWQRERASDVRKLISFLTHEDIMPRGKFLVVFLLSSVDDPRDPLIETF
                     CAFYQDLKGMENLLCICVPHFIFQGWKDLLEARLIKHQDEISSQCISALSLEEINGTI
                     LKLKSVTQGSKRLLPSIGLSTVLKKEEDIMTALEIICENECGTLLKEDKNKPLEFK
                     ASKEEDFYRGGKVSWNNFYFSSESYSSPFVRDKYERLEAMIQNCADSSKPTSTKIIH
                     LYHHPCGGTTLAMHILWELRKKFRCAVLKNKTVDFSEIGEQVTSLITYGAMNRQEYV
                     PVLLLVDDFEEQDNVYLLQYSIQTAIAKKYIRYEKPLVIILNCMRSONPEKSARIPDS
                     IAVIQQLSPKEQRAFELKLKEIKEQHKNFEDFYSFMIMKTNFNKEYIENVRNILKGQ
                     NIFTKEAKLFSFLALLNSYVPDTTISLSQCEKFLGIGNKKAFWGTEKFEDKMGTYSTI
                     LIKTEVISCGNYCGVRIHSLIAEFSLELKKSYHLNKSQIMLDMLTENLFFDTGMGK
                     SKFLQDMTLLLTHRDEHEGETGNWFSPFIEALHKDEGNEAVEAVLLESIHRFNPNA
                     FICQALARHFYIKKGDFGNALNWAKQAKIIEPDNSYISDTLGQVYKSKIRWWIEENGG
                     NGNISVDDLIALLDLAEHASSAFKESQQSEDREYEVKERLIYPKSKRRYDTYNIAGYQ
                     GEIEVGLYTIQIQLIPPFDNKNELSKRYMVNFVSGSSDIPGDPNNEYKLALKNYIPY
                     LTKLKFSLKKSFDFFDEYFVLLKPRNNIKQNEAKTRRKVAGYFKKYVDIFCLLEESQ
                     NNTGLGSKFSEPLQVERCRRNJVALKADKFSGLLEYLLKSQEDAISTMKCIVNEYTFL
                     LEQCTVKIQSKEKLNFILANIILSCIQTSRLVKPVEKLKDQLREVLQPIGLTYQFSE
                     PYFLASLLFWPENOQLDQHQSEQMKEYAQALKNSFKGQYKMHRTKQPIAYFLGKGKR
                     LERLVHKGKIDQCFKKTPDINSLWQSGDVWKEEKVQELLLRLQGRAENNCLYIEYGIN
                     EKTIPITPAFLGQLRSGRSIEKVSFYLGFSIGGPLAYDIEIV"
ORIGIN

Query Match   95.1%; Score 1472.4; DB 9; Length 6827;
Best Local Similarity 99.6%;
Matches 1476; Conservative   0; Mismatches   6; Indels   0; Gaps   0;

Qy   68 AGACTACTGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGG 127

```
Qy  1208 GCAGCAGAAGAAAATTCAGAGCAAAAACAAATAAAAAACAAATAAAAAAGAAGAGAGGGACCCAAAGTTG 1267
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1371 GCAGCAGAAGAAAATTCAGAGCAAAAACAAATAAAAAACAAATAAAAAAGAAGAGAGGGACCCAAAGTTG 1430

Qy  1268 GTTAAATTATTGACAGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATT 1327
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1431 GTTAAATTATTGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAGTACATT 1490

Qy  1328 CTTGTAACAAATAAATGCCACCGAGATCAAACAAACACTTAGATTTCCTGAAGGAAATT 1387
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1491 CTTGTAACAAATAAATGCCACCGAGATCAAACAAACACTTAGATTTCCTGAAGGAAATT 1550

Qy  1388 AAATGGTTTGCTGTATTGGAGTTTGATCCTGAGTTCTAACATCAATGGAGTGGTCAAAGCT 1447
         |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db  1551 AAATGGTTTGCTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCT 1610

Qy  1448 TACAAAGAAAGCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAAACC 1507
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1611 TACAAAGAAAGCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAAACC 1670

Qy  1508 ACACCAAATGAGACGATTTCTACTCTAAATCTTTACCATCAA 1549
         |||||||||||||||||||| |||||||||||||||||||||
Db  1671 ACACCAAATGAGACGATTTCACTCTAAATCTTTACCATCAA 1712
```

RESULT 4
AC000119/c
LOCUS   AC000119   191191 bp   DNA   linear   PRI 15-OCT-2003
DEFINITION Homo sapiens BAC clone CTB-104I4 from 7, complete sequence.
ACCESSION   AC000119
VERSION   AC000119.1 GI:1809226
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 191191)
AUTHORS   Sulston,J.E. and Wilson,R.
TITLE   Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE   99063792
PUBMED   9847074
REFERENCE   2 (bases 1 to 191191)
AUTHORS   Gibson,A. and Wamsley,P.
TITLE   The sequence of Homo sapiens BAC clone CTB-104I4
JOURNAL   Unpublished (2001)
REFERENCE   3 (bases 1 to 191191)
AUTHORS   Waterston,R.
TITLE   Direct Submission
JOURNAL   Submitted (31-JAN-1997)
REFERENCE   4 (bases 1 to 191191)
AUTHORS   Wilson,R.
TITLE   Direct Submission
JOURNAL   Submitted (15-OCT-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT   ----------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: sapiens@watson.wustl.edu
            ----------- Summary Statistics
            Center project name: H_RG104I04

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project and the Washington University Genome Sequencing Center. For
            additional information about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 or
            mailto:egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            Clone CTB-104I4 is from the first release of the human BAC library
            CITB-978SK-B. The library contains cloned DNA from the male
            fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad.
            Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8
            (1996). This clone is available from Research Genetics, Inc.
            (http://www.resgen.com).
            VBCTOR: pBeloBAC11
            Selection: chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of CTB-104I4.
            The actual end is at 191191 of CTB-104I4. The orientation of this clone
            is unknown.
FEATURES   Location/Qualifiers
     source   1..191191
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="CTB-104I4"
                     /clone_lib="CITB-978SK-B"
     repeat_region   1655..1726
                     /rpt_family="AcHobo"
     repeat_region   1752..1870
                     /rpt_family="Alu"
     repeat_region   1872..1960
                     /rpt_family="AcHobo"
     repeat_region   4418..4465
                     /rpt_family="CT-rich"
     repeat_region   5384..5471
                     /rpt_family="MIR"
     repeat_region   6010..6439
                     /rpt_family="MaLR"
     repeat_region   7812..7895
                     /rpt_family="MER1_type"
     repeat_region   7931..9288
                     /rpt_family="L1"
     repeat_region   11003..11945
                     /rpt_family="ERV1"
     repeat_region   12518..12710
                     /rpt_family="MIR"
     repeat_region   13031..13142
                     /rpt_family="L2"
     repeat_region   13985..14096
                     /rpt_family="MIR"
     repeat_region   14993..15226
                     /rpt_family="MIR"
     repeat_region   16163..16205
                     /rpt_family="AT_rich"
     repeat_region   16229..16266
                     /rpt_family="AT_rich"
     repeat_region   16354..16452
                     /rpt_family="L2"
     repeat_region   16505..16797
                     /rpt_family="Alu"
     repeat_region   16798..16870
                     /rpt_family="(TA)n"
     repeat_region   16901..17789

```
repeat_region    17802..18188    /rpt_family="L1"
repeat_region    18189..18219    /rpt_family="ERVL"
repeat_region    19807..20062    /rpt_family="(TCTG)n"
repeat_region    24317..24420    /rpt_family="Alu"
repeat_region    26099..26146    /rpt_family="(TA)n"
repeat_region    26468..26588    /rpt_family="L2"
repeat_region    26589..26784    /rpt_family="T2_type"
repeat_region    26792..26841    /rpt_family="Alu"
repeat_region    26845..27684    /rpt_family="T2_type"
repeat_region    27689..29178    /rpt_family="T2_type"
repeat_region    29183..29603    /rpt_family="ERV1"
repeat_region    30200..30243    /rpt_family="T2_type"
repeat_region    30249..30273    /rpt_family="(CAAAAA)n"
repeat_region    30274..30416    /rpt_family="AT_rich"
repeat_region    30521..30622    /rpt_family="L1"
repeat_region    30840..30970    /rpt_family="(TATATG)n"
repeat_region    31658..31711    /rpt_family="Alu"
repeat_region    31884..32110    /rpt_family="(CA)n"
misc_feature     34694..34782    /note="CpG_island (%GC=72.2, o/e=0.56, #CpGs=22)"
repeat_region    35891..36437    /rpt_family="MIR"
repeat_region    36645..37229    /rpt_family="L1"
repeat_region    37230..37275    /rpt_family="L1"
repeat_region    37276..38427    /rpt_family="(CAAAA)n"
repeat_region    38428..38699    /rpt_family="L1"
repeat_region    38700..39084    /rpt_family="Alu"
repeat_region    39085..39362    /rpt_family="L1"
repeat_region    39363..39600    /rpt_family="Alu"
repeat_region    39601..39885    /rpt_family="L1"
repeat_region    39986..40005    /rpt_family="L1"
repeat_region    40006..40343    /rpt_family="L1"
repeat_region    40344..40364    /rpt_family="MER1_type"
repeat_region    40381..40537    /rpt_family="L1"
repeat_region    40564..40976    /rpt_family="L1"
repeat_region    41007..41120    /rpt_family="L1"
repeat_region    41355..41630    /rpt_family="L1"
repeat_region    41965..42126    /rpt_family="Alu"

repeat_region    42148..42293    /rpt_family="L1"
repeat_region    42425..42532    /rpt_family="GA-rich"
repeat_region    42669..42699    /rpt_family="CR1"
repeat_region    42751..42939    /rpt_family="AT_rich"
repeat_region    42979..43272    /rpt_family="MIR"


Query Match    95.0%;  Score 1472;  DB 9;  Length 191191;
Best Local Similarity  99.7%;  Pred. No. 2.5e-268;
Matches 1475;  Conservative  0;  Mismatches  5;  Indels  0;  Gaps  0;

Qy    70 ACTACTGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGAC 129
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 94520 ACATATAGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGAC 94461

Qy   130 AAAAGAGGATGTAAATCAGTGGTTAGAAGTCATAAGATTGACCAAAAACACAGGGAAAT 189
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 94460 AAAAGAGATGTAAATCAGTGGTTAGAAGTCATAAGATTGACCAAAAACACAGGGAAAT 94401

Qy   190 TTTGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGT 249
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 94400 TTTGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTTGT 94341

Qy   250 TGATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 94340 TGATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCG 94281

Qy   310 GAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGAAGCCCAGTAAAAATGC 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 94280 GAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGAAGCCCAGTAAAAATGC 94221

Qy   370 TCCTAAAGACCAAACTGTCTCAAAAGGACGTAGAGAAACTTCAAAGCAAAAACAAAA 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 94220 TCCTAAAGACCAAACTGTCTCAAAAGGACGTAGAGAAACTTCAAAGCAAAAACAAAA 94161

Qy   430 GGGTAAAGAGAGAACCCAGATATGGCTAATCCGTCTGCAATGCAACTGCTAAGGTTC 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 94160 GGGTAAAGAGAGAACCCAGATATGGCTAATCCGTCTGCAATGCAACTGCTAAGGTTC 94101

Qy   490 TAAGTCACTAAAGTTGAGCTCATAGAAGATAAAATATGATTATACAAAGGAAAGGCAACC 549
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 94100 TAAGTCACTAAAGTTGAGCTCATAGAAGATAAAATATGATTATACAAAGGAAAGGCAACC 94041

Qy   550 ATCCATAGACCTGCACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTA 609
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 94040 ATCCATAGACCTGCACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTA 93981

Qy   610 CAAGTTGGATTTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACA 669
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 93980 CAAGTTGGATTTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACA 93921

Qy   670 TGAATTCAAAGCCTTCACAAATACAGCCACCACCAGAGGATGTCAAGATGAAATT 729
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 93920 TGAATTCAAAGCCTTCACAAATACAGCCACCACCAGAGGATGTCAAGATGAAATT 93861

Qy   730 TAGCAATGAGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTAT 789
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 93860 TAGCAATGAGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTAT 93801

Qy   790 TCATTTTGGAGTCAAAGACAAACCCATGGCATCAAATTGTTGGCATCAAATCAAAGTGA 849
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 93800 TCATTTTGGAGTCAAAGACAAACCCATGGCATCAAATTGTTGGCATCAAATCAAAGTGA 93741

Qy   850 TACCAAGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTGAAGACCA 909
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 93740 TACCAAGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTGAAGACCA 93681

Qy   910 TCAAGTCCAACAGCAAGAAGTGCATTCGAGCCAAGAGATTTGTGGAAGTTTTTACTGCC 969
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 93680 TCAAGTCCAACAGCAAGAAGTGCATTCGAGCCAAGAGATTTGTGGAAGTTTTTACTGCC 93621
```

```
Qy    970  AAATAGTAGTACTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGA 1029
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93620  AAATAGTAGTACTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGA 93561

Qy   1030  ATGCCAATATGATTACTTCCAGATTAAAATGCAAAATTACACACAACAAAATATGGGAACA 1089
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93560  ATGCCAATATGATTATTTCCAGATTAAAATGCAAAATTACACACAACAAAATATGGGAACA 93501

Qy   1090  AAGTAAAAAATCTCACTATTTGTGCCGAGATGGGACCAGCTCTAAGGACATTACGAAAAA 1149
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93500  AAGTAAAAAATCTCACTATTTGTGCCGAGATGGGACCAGCTCTAAGGACATTACGAAAAA 93441

Qy   1150  TAAAGTTGATTTCAGAGCGATTTAAAGCAGAGATTTAAAACACTGGCAGAGTCCAGAAAGC 1209
            ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db  93440  TAAAGTTGATTTCAGAGCGATTTAAAGCAGATTTAAAACACTGGCAGAGTCCAGAAAAGC 93381

Qy   1210  AGCAGAGAAAAATTCAGAGCAAAAACAAATAAAAAAGAAAGAGAGGGACCAAAGTTGGT 1269
            |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db  93380  AGCAGAGAAGAAATTCAGAGCAAAAACAAATAAAAAAGAAAGAGAGGGACCCAAAGTTGGT 93321

Qy   1270  TAAATTATTGACAGGAAATCAAGATTTGTTAGATAATTCATACTACTAACAGTACATTCT 1329
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93320  TAAATTATTGACAGGAAATCAAGATTTGTTAGATAATTCATACTATCATGAACAGTACATTCT 93261

Qy   1330  TGTAACAAATAAATGCCACCCAGATCAAACAAAACACTTAGATTTCCTGAGGAAATTAA 1389
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93260  TGTAACAAATAAATGCCACCCAGATCAAACAAAACACTTAGATTTCCTGAGGAAATTAA 93201

Qy   1390  ATGGTTTGCTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCTTA 1449
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93200  ATGGTTTGTGTATTGGAGTTTGATCCTGAGTCTTAACATCAATGGAGTGGTCAAAGCTTA 93141

Qy   1450  CAAAGAAAGCCGAGTAGCAAACCTTCACTTTCCAAGTGTATATGTAGAACAGAAAACCAC 1509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  93140  CAAAGAAAGCCGAGTAGCAAACCTTCACTTTCCAAGTGTATATGTAGAACAGAAAACCAC 93081

Qy   1510  ACCAAATGAGACAGATTTCTCACTCTAAATCTTTACCATCAA 1549
            |||||||||||||||||||||||||||||||||||||||||||
Db  93080  ACCAAATGAGACAGATTTCTCACTCTAAATCTTTACCATCAA 93041
```

RESULT 5
AC092871/c
LOCUS   AC092871   172371 bp   DNA   linear   HTG 12-JUN-2002
DEFINITION Pan troglodytes clone RP43-120E8, WORKING DRAFT SEQUENCE, 3 ordered
            pieces.
ACCESSION   AC092871
VERSION   AC092871.2 GI:21392483
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE   Pan troglodytes (chimpanzee)
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 172371)
AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
TITLE   NISC Comparative Sequencing Initiative
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 172371)
AUTHORS   Green,E.D.
TITLE   Direct Submission
JOURNAL   Submitted (01-AUG-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3 (bases 1 to 172371)

AUTHORS   Green,E.D.
TITLE   Direct Submission
JOURNAL   Submitted (12-JUN-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
            On Jun 12, 2002 this sequence version replaced gi:15055304.
COMMENT   -------------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
            Center project name: ank
            Center clone name: 120E08

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassembles, the low-quality sequence ends of sequence
            contig have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; version 0.990319
            Assembly program: Phrap; version 0.990319
            Consensus quality: 171808 bases at least Q40
            Consensus quality: 172035 bases at least Q30
            Consensus quality: 172123 bases at least Q20
            Insert size: 129000; agarose-fp
            Insert size: 137000; pulse-field-gel
            Insert size: 172371; sum-of-contigs
            Quality coverage: 15.17x in Q20 bases; agarose-fp
            Quality coverage: 14.29x in Q20 bases; pulse-field-gel
            Quality coverage: 11.37x in Q20 bases; sum-of-contigs
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *   1   77640: contig of 77640 bp in length
            *   77640   77740: gap of unknown length
            *   77641   77740: contig of 77640 bp in length
            *   77741 102026: contig of 24286 bp in length
            *   102007 102126: gap of unknown length
            *   102127 172371: contig of 70245 bp in length
FEATURES   Location/Qualifiers
     source   1..172371
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-120E8"
                     /clone_lib="RP43"
     misc_feature   1..77640
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature   1..3202
                     /note="clone overlaps with GenBank Accession Number
                     AC098646 clone RP43-44G6 (center project name anj)"
     misc_feature   77741..102026
                     /note="assembly_fragment"
     misc_feature   102127..172371
                     /note="assembly_fragment
                     clone_end:T7
```

ORIGIN

    vector_side:right"

Query Match   94.5%; Score 1464; DB 2; Length 172371;
Best Local Similarity   99.3%; Pred. No. 8.4e-267;
Matches 1470; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy      70 ACTACTGGGTATCAGAAGCGAAGCAACTAACCTTCCAGAAAATACAGATGATTGGAC 129
           | ||| || |||||||||||||||||||||||||||||||||||||||||||||||
Db  117802 ACATATAGGTATCAGAAGCAACTAACCTTCCAGAAAATACAGATGATTGGAC 117743

Qy     130 AAAAGAAGAGATGTAAATCAGTGGTTAGAAAGTCATAAGAACACATCTTGGT 189
           |||||||||| ||||||||||||||||||||||| |||||||||||||||||
Db  117742 AAAAGAAGAGGTAAATCAGTGGTTAGAAAGTGGTTAAAAAAGAACACATCTTGT 117683

Qy     190 TTTGACTGAACAAGAGGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTTGT 249
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117682 TTTGACTGAACAAGAGGAGCAGTCTTGAAGTGGTTAAAAAAAGAACATCTTGT 117623

Qy     250 TGATATGGGCATCACACATGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCG 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117622 TGATATGGGCATCACACATGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCG 117563

Qy     310 GAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGC 369
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117562 GAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGC 117503

Qy     370 TCCTAAAGACCAAACTGTCTCAAAAGGACGTAGAGAAACTTCAAAGCAAAAACAAAA 429
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117502 TCCTAAAGACCAAACTGTCTCAAAAGGACGTAGAGAAACTTCAAAGCAAAAGACAAAA 117443

Qy     430 GGGTAAAGAGACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTC 489
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117442 GGGTAAAGAGACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTC 117383

Qy     490 TAAGTCACTAAAAGTTGAGCTCATAGAAGATAAAAATAGATAAAAATAGATTAATCGTTCAAAGGAAAGGCAACC 549
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117382 TAAGTCACTAAAAGTTGAGCTCATAGAAGATAAAAATAGATTAATCAGTTA 117323

Qy     550 ATCCATAGACCTGACATGTGTATCATATCATTCATTTGATGAATTCAGTTCGTTA 609
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117322 ATCCATAGACCTGACATGTGTATCATATCATTCATTTGATGAATTCAGTTCGTTA 117263

Qy     610 CAAGTTGGATTTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCCGATACA 669
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117262 CAAGTTGGATTTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCCGATACA 117203

Qy     670 TGAATTCAAAGCCTTCCACAAATAGCAGACAGCCAACAGCCACAGAAGAGGATGTCAAGATGAAATT 729
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117202 TGAATTCAAAGCCTTCCACAAATAGCAGACAGCCAACAGCCACAGAAGAGGATGTCAAGATGAAATT 117143

Qy     730 TAGCAATGAGGTTTCCGATTGCTTCAGCTTGTGATGAATTCAGTGATGAATTCACGTACCGACTAT 789
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117142 TAGCAATGAGGTTTCCGATTGCTTCAGCTTGTGATGAATTCACGGCACCAATGGCGACTAT 117083

Qy     790 TCATTTTGGAGTCAAAGACAAACAAACCCCATGGGAAATTGTTGGCATCAAAGTCACCAATGA 849
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117082 TCATTTTGGAGTCAAAGACAAACAAACCCCATGGGAAATTGTTGGCATCAAAGTCACCAATGA 117023

Qy     850 TACCAAGGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCA 909
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117022 TACCAAGGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCA 116963

Qy     910 TCAAGTCCAACAAGCAAAGAGAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCC 969
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116962 TCAAGTCCAACAAGCAAAGAGAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCC 116903

Qy     970 AAATAGTACTCTATCTGACAGATTGTTATTGAATGGAGACATTATTCCACAGTTCTCTGA 1029
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116902 AAATAGTACTCTATCTGACAGATTTGTTATTGGACACTTATTCCACAGTTCTCTGA 116843

Qy    1030 ATGCCAATATGATTATTCCAGATTAAAATGCAAAATCAAATTACAACAACAAAATATGGGAACA 1089
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  116842 ATGCCAATATGATTATTCCAGATTAAAATGCAAAATCAAATTACAACAACAAAACATGGGAACA 116783

Qy    1090 AAGTAAAAAATTCACTATTTGTGCGAGATGGACCAGCTCTAAGGACATTACGAAAAA 1149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116782 AAGTAAAAAATTCACTATTTGTGCGAGATGGACCAGCTCTAAGGACATTACGAAAAA 116723

Qy    1209 TAAAGTTGATTTCAGAGCATTTAAAGCAGATTTTAAACACTGGCAGAGTCCAGAAAAGC 1209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116722 TAAAGTTGATTTCAGAGCATTTAAAGCAGATTTTAAACACTGGCAGAGTCCAGAAAAGC 116663

Qy    1269 AGCAGAGAAAATTCAGAGCAAAACAAATAAAAAAGAAAGAAGAGGGACCAAAGTTGGT 1269
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116662 AGCAGAGAAAATTCAGAGCAAAACAAATAAAAAAGAAAGAAGAGGGACCAAAGTTGGT 116603

Qy    1329 TAAATTATTGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACTACATTCT 1329
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116602 TAAATTATTGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAACTACATTCT 116543

Qy    1389 TGTAACAAATAAATGCCACCCGAGTTGATCCAAGATCAAACAAAACACTTAGATTTCCTGAGGAAATTAA 1389
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116542 TGTAACAAATAAATGCCACCCGAGTTGATCCAAGATCAAACAAAACACTTAGATTTCCTGAGGAAATTAA 116483

Qy    1449 ATGGTTGCTGATTGGAGTTGATCCTGAGTCTAACATCAATGAGTGGTCAACATGGTCAAAGCTTA 1449
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116482 ATGGTTGCTGATTGGAGTTGATCCTGAGTCTAACATCAATGAGTGGTCAACATGGTCAAAGCTTA 116423

Qy    1509 CAAAGAAAGCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTGAACAGAAAAACCAC 1509
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116422 CAAAGAAAGCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTGAACAGAAAAACCAC 116363

Qy    1549 ACCAAATGAGACGATTTCTACTCTAAATCTTTACCATCAA 1549
           ||||||||||||||||||||||||||||||||||||||||
Db  116362 ACCAAATGAGACGATTTCTACTCTAAATCTTTACCATCAA 116323
```

Query Match   94.4%; Score 1462.4; DB 6; Length 4771;
Best Local Similarity   99.9%; Pred. No. 3.5e-266;
Matches 1463; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      86 ATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGGATGTAAAT 145
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGGATGTAAAT 60

Qy     146 CAGTGGTTAGAAGTGTCATAAGATGACAAGGGAAATTTGACTGACAACAAGAC 205
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CAGTGGTTAGAAGTGTCATAAGATGACAAGGGAAATTTGACTGACAACAAGAC 120

Qy     206 GTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA 265
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA 180

Qy     266 CATGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAGCCATTGAA 325
```

```
Db   181 CATGGACCAGCTATTCAAATAGAGAGAACTATTCAAAGAGAATTGCCGAAAAATGCTCCTAAAGACCAAACT 240
                                                                              ...
Qy   326 GATTCGATTCAGACACATCTCAAGAATGGGAAAGCCCAGTAAAAAATGCTCCTAAAGACCAAACT 385
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GATTCGATTCAGACATCAAGAATGGGAAAGCCCAGTAAAAAATGCTCCTAAAGACCAAACT 300

Qy   386 GTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGGGTAAAGAACCCA 445
             ||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGGGTAAAGAACCCA 360

Qy   446 GATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCAAGTCACTAAAAGTT 505
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCAAGTCACTAAAAGTT 420

Qy   506 GAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA 565
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACA 480

Qy   566 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTTAGT 625
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTTAGT 540

Qy   626 CTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCAAAGCCTTC 685
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCAAAGCCTTC 600

Qy   686 ACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAATTAGCAATGAGGTTTTC 745
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ACAAATACAGCAACAGCCACAGAAGAGGATGTCAAGATGAAATTAGCAATGAGGTTTTC 660

Qy   746 CGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAAA 805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAAA 720

Qy   806 GACAAACCCATGGGAAAAATTGTTGGCATCAAAGTCACCAAATGATACCAAGGAAGCCCTC 865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GACAAACCCATGGGAAAAATTGTTGGCATCAAAGTCACCAAATGATACCAAGGAAGCCCTC 780

Qy   866 ATTAACCATTTCAATCGATGATAAACAAGTATTTTGAAGACCATCCAAGTCCAACAGCA 925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 ATTAACCATTTCAATCGATGATAAACAAGTATTTTGAAGACCATCCAAGTCCAACAAGCA 840

Qy   926 AAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTTACTCGCCAAATAGTACTCTATCT 985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AAGAAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTTACTCGCCAAATAGTACTCTATCT 900

Qy   986 GACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAATATGATTAC 1045
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAATATGATTAT 960

Qy  1046 TTCCAGATTAAAATGCAAAATGCAAAATTACAACACAAAGACAAAGTAAAAAATTCTCA 1105
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TTCCAGATTAAAATGCAAAATGCAAAATTACAACACAAACAAAGACAAAGTAAAAATTCTCA 1020

Qy  1106 CTATTTGTGCGAGATGGGACCAGCTCTCAAGGACATTACGAACGTTGATTTGATTTCAGA 1165
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CTATTTGTGCGAGATGGGACCAGCTCTCAAGGACATTACGAACGTTGATTTGATTTCAGA 1080

Qy  1166 GCATTAAAGCAGATTTTAAAACACTGGCAGTGGGAACCAAGTCAGCAGAAGAAAAATTC 1225
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GCATTAAAGCAGATTTTAAAACACTGGCAGTGGGAACCAAGTCAGCAGAAGAAAAATTC 1140

Qy  1226 AGAGCAAAAACAAAATAAAAGAAGAGAGAGAGGACCAAAGTTGGTTAAATTATTGACAGGA 1285
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 AGAGCAAAAACAAAATAAAAGAAGAGAGAGAGGACCAAAGTTGGTTAAATTATTGACAGGA 1200

Qy  1286 AATCAAGATTTGTTAGATAATTCATCATGAACAGTACATTCTTGTAACAAATAAATGC 1345
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 AATCAAGATTTGTTAGATAATTCATCATGAACAGTACATTCTTGTAACAAATAAATGC 1260

Qy  1346 CACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAAAATTAAATGTTTTGCTGTATTG 1405
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAAAATTAAATGTTTTGCTGTATTG 1320
```

```
Qy  1406 GAGTTTGATCCTGAGTCTCAACATCAATGAGTGGTCAAAGCTTACAAAGAAAGCCGAGTA 1465
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 GAGTTTGATCCTGAGTCTCAACATCAATGAGTGGTCAAAGCTTACAAAGAAAGCCGAGTA 1380

Qy  1466 GCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAACCACACCAAATGAGACGATT 1525
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 GCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAACCACACCAAATGAGACGATT 1440

Qy  1526 TCTACTCTAAATCTTACCATCAA 1549
             |||||||||||||||||||||||
Db  1441 TCTACTCTAAATCTTACCATCAA 1464
```

RESULT 7

```
LOCUS        AX098228                6233 bp    DNA        linear    PAT 30-MAR-2001
DEFINITION   Sequence 3 from Patent WO0118208.
ACCESSION    AX098228
VERSION      AX098228.1  GI:13515357
KEYWORDS     .
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1
  AUTHORS    Peyman,J.A., da Silva,A., Hochman,P. and Hsu.A.
  TITLE      Interferon induced polynucleotides and proteins encoded thereby
  JOURNAL    Patent: WO 0118208-A 3 15-MAR-2001;
             Curagen Corporation (US); BIOGEN, INC. (US)
FEATURES             Location/Qualifiers
     source          1..6233
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             1..4770
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC35438.1"
                     /db_xref="GI:13515358"
                     /db_xref="REMTREMBL:CAC35438"
                     /translation="MAKQLNLPENTDDWTKEDVNQWLESHKIDQKHREILTEQDVNGA
                     VLKWLKEHLVDMGITHGPAIQEEELPEKLRKTAIEDSIQTSKMGKPSKNAPKDQTVS
                     QKERREISKQKQKGKENPDMANPSAMSTTAKGSKSLKVELIEDKIDYTKEROPSIDLT
                     CVSYPFDEFSNPYRYKLDFSLQPETGPGNLIDFHEFKAFTNTATATEDVKMKFSNE
                     VFRFASACMNSRTNGTIHFGVADKPHGKIVGIKVTNDTKEALINHFNLMINKYFEDHQ
                     VQQAKKCIREPRFVEVLLPNSTLSDRFVIEVDIIPQFSECQYDYFQIKMQNYNNKIWE
                     QSKKFSLFVRDGTSSKDITKNKVDFRAFKADFKTLAESRKAAEKFRAKTNKKEREGP
                     KLVKLLTGNQDLLDNSYEQYILVTNKCHPDQTKHLDFLKEIKWFAVLBFDPSSNING
                     VVKAYKSRVANLHFPSVYVEQKTTPNETISTLNLYHQPSWIFCNGRLDLDSEKYKPF
                     DPSSWQRERASDVRKLISFLTHEDIMPRGKFLVFLLLSVDDPRDPLIETFCAFYQD
                     LKGMENILCICVPHIFQGWKDLLEARLIKHQDEISSQCISALSLEEINGTILKLKSV
                     TQSSKRLLPSIGLSTVLLKKEEDIMTALEIICENECEGTLLEKDKNKPLEFKASKEED
                     FYRGGKVSWWNFYFSSESYSSPFVKRDYKYERLEAMIONCADSSKPTSTKIIHLYHHPG
                     CGGTTLAMHILWELRKKFRCAVLKNKTVDFSEIGEQVTSLIITYGAMNRQEYVPVLLV
                     DDFEEQDNVYLLQYSIQOTAIAKKYIRYEKPLVIILNCMRSQNPEKSARIPDSIAVIQQ
                     LSPKEQRAPELKLKEIKEQHKNFEDFVSFMIMKTNFNKEYIENVVRNILKGQNIFTKEV
                     AKLFSFLAILNSYVPDTTISLSQCEKFIGIGNKKAFWGTEKFEDKMGTYSTILLKTEV
                     IECGNYCGVRIIHSLIAEFSLREELKKSYHLNKSQIMLDMLTENLFFDTGMGKSKFLQD
                     MHTLLTRHRDEHEGETGNWFSPFIEALHKDEGNEAVAVLLESIHRENAFICQAL
                     ARHFYIKKKDFGNALNWAKQAKIIEPDNSYISDTLGQVYKSKIRWWIRENGGNGNISV
                     DDLIALLDLAEHRASSAFKESQQQSEDREYEVEKRTRRKVAGYFKKYVDIFCLLEESQNNTGLG
                     LYTIQILOLIPFFDNKNELSKRYMVNFVSGSSDIPGDPNNEYKLALKNYIPYLTKLKF
                     SLKKSPDFFDEYFLLKPRNNIKQNEEAKTRRKVAGYFKKYVDIFCLLEESQNNTGLG
                     SKFSEPLQVERCRRNIVALKADKFSGLLEYLIIKSQEDAISTMKCIVNEYTFLLEQCTV
                     KIQSKEKLNFILANIILSCIQPTSRLVFKVEKLKDQLREVLQPIGLTYQFSEPYFLVAS
                     KGKIDQCFKKTPDINSLWQSGDVWKEBKVQELLLRLQGRAENNCLYIEYGINEKITIP
                     ITPAFLGQLRSGRSIEKVSFYLGFSIGGPLAYDIEIV"
```

ORIGIN

```
Query Match                94.4%;   Score 1462.4;  DB 6;  Length 6233;
Best Local Similarity      99.9%;   Pred. No. 3.3e-266;
```

Matches 1463; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        86 ATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGGATGTAAAT 145
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAGGATGTAAAT 60

Qy       146 CAGTGGTTAGAAGTCATAGAGATTGACCAAAAACACAGGGAAATTTTGACTGAACAGAC 205
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAGTGGTTAGAAGTCATAGAGATTGACCAAAAACACAGGGAAATTTTGACTGAACAGAC 120

Qy       206 GTGAATGGAGCAGTCTTGAAGTGGTTAAAAAGAACATCTTCTTGTGATATGGGCATCACA 265
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GTGAATGGAGCAGTCTTGAAGTGGTTAAAAAGAACATCTTCTTGTGATATGGGCATCACA 180

Qy       266 CATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAGCCATTGAA 325
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 CATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAACAGCCATTGAA 240

Qy       326 GATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAAAAATGCTCTAAAGACCAAACT 385
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAAAAATGCTCTAAAGACCAAACT 300

Qy       386 GTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAAACAAAAGGGTAAAGAGAACCCA 445
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GTGTCTCAAAAGGAACGTAGAGAACTTCAAAGCAAAAACAAAAGGGTAAAGAGAACCCA 360

Qy       446 GATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTT 505
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTT 420

Qy       506 GAGCTCATAGAAGATAAAATAGATATTACAAAGGAAAGGCAACCATCCATAGACCTGACA 565
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 GAGCTTCATAGAAGATAAAATAGATATTACAAAGGAAAGGCAACCATCCATAGACCTGACA 480

Qy       566 TGTGTATCATATCCATTTGATGAATTCAGTGAATCCATATCGTTACAAGTTGATTTTAGT 625
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 TGTGTATCATATCCATTTGATGAATTCAGTGAATCCATATCGTTACAAGTTGATTTTAGT 540

Qy       626 CTACAGCCTGAAACACAGGACCAGGCAATCTCATTGATCCGATAACATGAATTCAAAGCCTTC 685
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 CTACAGCCTGAAACACAGGACCAGGCAATCTCATTGATCCGATAACATGAATTCAAAGCCTTC 600

Qy       686 ACAAATACAGCAACAGCCCAGAGAGATGTCAAGATGAAATTTAGCAATGAGAGGTTTTC 745
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 ACAAATACAGCAACAGCCCAGAGAGATGTCAAGATGAAATTTAGCAATGAGAGGTTTTC 660

Qy       746 CGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAAA 805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTTGGAGTCAAA 720

Qy       806 GACAAACCCCATGGAAAAATTGTTGGCATCAAAGTCACCAAAGTACCAAGGAAGCCCCTC 865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GACAAACCCCATGGAAAAATTGTTGGCATCAAAGTCACCAAAGTACCAAGGAAGCCCCTC 780

Qy       866 ATTAACCATTTCATCTGATGATAAACAAGTATTTGAAGACCATCCAACAAGTCCAACAAGCA 925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 ATTAACCATTTCATCTGATGATAAACAAGTATTTGAAGACCATCCAACAAGTCCAACAAGCA 840

Qy       926 AAGAAGTGCATTCGAGAGCGCCAAGAGTTTGTGTGGAAGTTTTACTGCCAAATAGTACTTCTATCT 985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 AAGAAGTGCATTCGAGAGCGCCAAGAGTTTGTGTGGAAGTTTTACTGCCAAATAGTACTTCTATCT 900

Qy       986 GACAGATTTGTTATTGGAAGTGGACATTATTCCACAGTTCCTGAAATGCCAATATGATTAC 1045
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 GACAGATTTGTTATTGGAAGTGGACATTATTCCACAGTTCCTGAATGCCAATATGATTAT 960

Qy      1046 TTCCAGATTAAAATGCAAAATTACAACAACAAATATGGAACAAAATAAAAATTCTCA 1105
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 TTCCAGATTAAAATGCAAAATTACAACAACAAATATGGAACAAAATAAAAAATTCTCA 1020

Qy      1106 CTATTTGTGCGAGATGGGACCAGCTCTTAAGGACATTACGAAAAAATAAAGTTGATTTCAGA 1165
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 CTATTTGTGCGAGATGGGACCAGCTCTTAAGGACATTACGAAAAAATAAAGTTGATTTCAGA 1080
```

```
Qy      1166 GCATTTAAAGCAGATTTAAAACACTGGCAGAGTCCAGAGTCCAGAAAAGCAGCAAGAAAAATTC 1225
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 GCATTTAAAGCAGATTTAAAACACTGGCAGAGTCCAGAGTCCAGAAAAGCAGCAAGAAAAATTC 1140

Qy      1226 AGAGCAAAAACAAATAAAAAAGAAAAGAAGAAGAAGAGGGGACCAAAGTTGGTTAAATTATTGACAGGA 1285
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 AGAGCAAAAACAAATAAAAAAGAAAAGAAGAAGAAGAGGGGACCAAAGTTGGTTAAATTATTGACAGGA 1200

Qy      1286 AATCAAGATTTGTTAGATAATTCATACATGAACAGTACATTCTGTGAACAATAAATGC 1345
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 AATCAAGATTTGTTAGATAATTCATACATGAACAGTACATTCTGTGAACAATAAATGC 1260

Qy      1346 CACCCAGATCAAACAAAAACACTTAGATTCCTGAAGGAAATTAAAGGTTGCTGTATTG 1405
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 CACCCAGATCAACAAACAAAACACTTAGATTCCTGAAGGAAATTAAAGGTTTGCTGTATTG 1320

Qy      1406 GAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCTTACAAAGAAAGCCAGTA 1465
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 GAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCTTCACAAGAAAGCCCAGTA 1380

Qy      1466 GCAAACCTTCACTTTCCCAAGTGTATATGTGAGACAAGAAAACCACCACCAAATGAGACGATT 1525
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 GCAAACCTTCACTTTCCCAAGTGTATATGTGAGACAAGAAAACCACCACCAAATGAGACGATT 1440

Qy      1526 TCTACTCTAAATCTTTACCATCAA 1549
             |||||||||||||||||||||||
Db      1441 TCTACTCTAAATCTTTACCATCAA 1464
```

RESULT 8
LOCUS   AC092513/c
DEFINITION AC092513   176350 bp   DNA   linear   HTG 19-JUN-2002
            Papio anubis clone RP41-474E21, WORKING DRAFT SEQUENCE.
ACCESSION   AC092513
VERSION   AC092513.2 GI:21465392
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE   Papio anubis (olive baboon)
  ORGANISM Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1 (bases 1 to 176350)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Pagujirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE   NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 176350)
  AUTHORS   Green,E.D.
  TITLE   Direct Submission
  JOURNAL   Submitted (13-JUL-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3 (bases 1 to 176350)
  AUTHORS   Green,E.D.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-JUN-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT   On Jun 19, 2002 this sequence version replaced gi:14717326.
            --------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_zoo@nhgri.nih.gov
            --------------- Project Information
                Center project name: aml

Center clone name: 474E21

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclone, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8X average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.

--------- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 176118 bases at least Q40
Consensus quality: 176233 bases at least Q30
Consensus quality: 176319 bases at least Q20
Insert size: 135000; agarose-fp
Insert size: 172000; pulse-field-gel
Insert size: 176350; sum-of-contigs
Quality coverage: 17.59x in Q20 bases; agarose-fp
Quality coverage: 13.81x in Q20 bases; pulse-field-gel
Quality coverage: 13.47x in Q20 bases; sum-of-contigs

----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1   176350: contig of 176350 bp in length.

FEATURES   Location/Qualifiers
source   1..176350
                /organism="Papio anubis"
                /mol_type="genomic DNA"
                /db_xref="taxon:9555"
                /clone="RP41-474E21"
                /clone_lib="RP41"
misc_feature   1..176350
                /note="assembly_fragment
                clone end:SP6
                vector side:left
                clone end:T7
                vector side:right"
misc_feature   1..44788
                /note="clone overlaps with GenBank Accession Number
                AC092407 clone RP41-402M16 (center project name amk)"
misc_feature   136968..176350
                /note="clone overlaps with GenBank Accession Number
                AC092515 clone RP41-6601 (center project name ami)"

ORIGIN

Query Match   90.2%;   Score 1396.8;   DB 2;   Length 176350;
Best Local Similarity 96.5%;   Pred. No. 4e-254;
Matches 1428;   Conservative 0;   Mismatches 52;   Indels 0;   Gaps 0;

Qy   70 ACTACTGGGTATCAGAATGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGAC 129
         || || |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 99493 ACATGTAGGTATCAGAATGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGAC 99434

Qy   130 AAAGAGGATGTAAATCCGTGGTTAGAAAGTCATCAGAATCAGAAGTGACCAAAAACACAGGGAAAT 189
         |||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||||||||
Db 99433 CAAAGAGGATGTAAATCGGTGGTTAGAAAGTCACAAGAATCAGAAGTGACCAAAAACACAGGGAAAT 99374

Qy   190 TTTGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGT 249
         |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 99373 TTTGACTGGACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGT 99314

Qy   250 TGATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACATTCAAAGAATTGCG 309
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 99313 TGATATGGGCATCACACATGGACCAGCTATTCAAATAGAAGAACATTCAAAGAATTGCA 99254

Qy   310 GAAACAGCCATTGAAGATTCGATTCAGACATCAAGCATCTAAGATGGGAAAGCCCAGTAAAAATGC 369
         ||||||||||||||||||||| ||||||||| |||||||||||||||||||||||||
Db 99253 GAAACAGCCATTGAAGATTCAATTCAGACATCTAAGATGAGAAAGCCCAGTAAAAATGC 99194

Qy   370 TCCTAAAGACCAAACTGTGTCTCAAAAGAACGTAGAGAAACTTCAAAGCAAAAACAAAA 429
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 99193 TCCTAAAGACCAAACCGTGTCTCAAAAGAACGTAGAGAAACTTCAAAGCAAAAACAAAA 99134

Qy   430 GGGTAAAGAGAGACCCAGATATGGCTAATCCGTCGCAATGAGTCAATGAGTACAACTGC 489
         ||||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||
Db 99133 GGATAAAGAGAGACCCAGATATGGCTAATCCATCGCAATCATCACAATGAGTACAATTG 99074

Qy   490 TAAGTCACTAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACC 549
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 99073 TGAGTCACTAAAGTTGAGCTCATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACC 99014

Qy   550 ATCCATAGACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTA 609
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 99013 ATCCATAGACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTA 98954

Qy   610 CAAGTTGGATTTAGTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCGATACA 669
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98953 CAAGTTGGATTTATTCTACAGCCTGAAACAGGACCAGGCAATCTCATTGATCCAATACA 98894

Qy   670 TGAATTCAAAGCCTTCACAAATACAGCAACAGCCACAGAGGATGTCAAGATGAAAATT 729
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98893 TGAATTCAAAGCCTTCACAAATACAGCAAATACAGCCACAGAGGATGTTAAGATGAAATT 98834

Qy   730 TAGCAATGAGGTTTCCGATTGCTTCCAGCTTGTGATGAATTCAGTACCAATGGCACTAT 789
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98833 TAGCAATGAGGTTTCCGATTGCTTCCAGCTTGTGATGAATTCAGTACCAATGGCACTAT 98774

Qy   790 TCATTTTGGAGTCAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGA 849
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98773 TCATTTTGGAGTCAAGACAAACAACCCATGGGAAAATTGTTGGCGTCAAAGTCACCAGTGA 98714

Qy   850 TACCAAGGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAACAGTATTTGAAGACCA 909
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98713 TACCAAGGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAACAGTATTTTGAAGACCA 98654

Qy   910 TCAAGTCCAACAGCAAAGAGTGCATTCGAGAGCCAAGATTTGTGTGGAAGTTTTACTGCC 969
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98653 TCAGGTCCAACAAGCAAAGAGTGCATTCGAGAGCCAAGATTTGTGTGGAGGTTTTACTGCC 98594

Qy   970 AAATAGTACTCTATCTGACAGATTGTTATTGACAGATTTATTGAAGTGGACATTATTG 1029
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98593 AAATAGTACTTTATCTGACAGATTGTTATTGACAGATTGGAAGTGGACATTATTG 98534

Qy 1030 ATGCCAATATGATTACTTCCAGATTAAAATGCAAAATTCAAAATTACAACAAAAATATGGGAACA 1089
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98533 ATGCCAATATGATTACTTCCAGATTAAAATGCAAAATTCAAAATTACAACAACAAAACATGGGAACA 98474

Qy 1090 AGTAAAAATTCTCACTATTTGTGCGAGATGGACCAGCTTCAAGGACATTACGAAAAA 1149
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98473 AGTAAAAAATTCTCACTATTTGTGCGAGATGGACCAGCTTCAAGGACATTATGAAAAA 98414

Qy 1150 TAAAGTTGATTTCAGAGCATTTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAAAGC 1209
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98413 TAAAGCTGATTTCAGAGAACATTTAAAGCAGATTTTAAAACACTGGCAGAGTCCAGAAAAGC 98354

Qy 1210 AGCAGAAGAAATTCAGAGACAAAAACAAATAAAAGAAGAAAGAGAGGGACCAAAGTTGGT 1269
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98353 AGCAGAAGAAATTCAGAGACAAAAACAAATAAAAGAAGAAAGAGGGGACCAAAGTTGGT 98294

Qy 1270 TAAATTATTGACACGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAAGTACATTCT 1329
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98293 TAAATTATTGACACGGAAATCAAGATTTGTTAGATAATTCATACTATGAACAAGTACATTCT 98234

```
Qy      1330  TGTAACAAATAAATGCCACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAAATTAA  1389
              ||||||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     98233  TGTAACGAATAAATGTCACCCAGATCAAACAAAACACTTAGATTTCCTGAAGGAAATTAA  98174

Qy      1390  ATGGTTTGCTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCTTA  1449
              |||||||||| ||||| ||||||||||||||||||||||||||||||||| ||| |||
Db     98173  ATGGTTTGCTGTGTTGGAGTTCGATCCTGAGTCTAACATCAATGGAGTGGTCCAAGATTA  98114

Qy      1450  CAAAGAAAGCCGAGTAGCAAACCTTCACTTCCAAGTGTATATGTAGAACAGAAAACCAC   1509
              |||| ||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     98113  CAAAGCAAGCCGAGTAGCAAACCTTCACTTCCAAGTATATGTAGAAGAGGAAAACCAC    98054

Qy      1510  ACCAAATGAGACGATTTCTACTCTAAATCTTACCATCAA  1549
              ||||||||||||||||||||||||| ||||||||||||||
Db     98053  ACCAAATGAGAAGATTTCTACTCTGAATCTTACCATCAA  98014
```

```
RESULT 9
AC105435/c
LOCUS       AC105435             175424 bp    DNA     linear   HTG 18-JUL-2002
DEFINITION  Sus scrofa clone RP44-307B12, WORKING DRAFT SEQUENCE, 5 ordered
            pieces.
ACCESSION   AC105435
VERSION     AC105435.2  GI:21903575
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Sus scrofa (pig)
ORGANISM    Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 175424)
AUTHORS     Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
TITLE       NISC Comparative Sequencing Initiative
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 175424)
AUTHORS     Green,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (05-JAN-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 175424)
AUTHORS     Green,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (18-JUL-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
            On Jul 18, 2002 this sequence version replaced gi:18071331.
COMMENT     ------------ Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_zoo@nhgri.nih.gov
            ------------ Project Information
            Center project name: caw
            Center clone name: 307B12
            ------------
            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ------------ Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 173481 bases at least Q40
            Consensus quality: 174362 bases at least Q30
            Consensus quality: 174801 bases at least Q20
            Insert size: 175000; agarose-fp
            Insert size: 175024; sum-of-contigs
            Quality coverage: 11.12x in Q20 bases; agarose-fp
            Quality coverage: 9.66x in Q20 bases; sum-of-contigs
            ------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *    1       11441: contig of 11441 bp in length
            *    11442   11541: gap of unknown length
            *    11542   23012: contig of 11471 bp in length
            *    23013   23112: gap of unknown length
            *    23113   31457: contig of 8345 bp in length
            *    31458   31557: gap of unknown length
            *    31558  148674: contig of 117117 bp in length
            *   148675  148774: gap of unknown length
            *   148775  175424: contig of 26650 bp in length.
FEATURES             Location/Qualifiers
     source          1..175424
                     /organism="Sus scrofa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9823"
                     /clone="RP44-307B12"
                     /clone_lib="RP44"
     misc_feature    1..36234
                     /note="clone overlaps with GenBank Accession Number
                     AC105370 clone RP44-36L5 (center project name cav)"
     misc_feature    1..11441
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    11542..23012
                     /note="assembly_fragment"
     misc_feature    23113..31457
                     /note="assembly_fragment"
     misc_feature    31558..148674
                     /note="assembly_fragment"
     misc_feature    148775..175424
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN
```

```
Query Match         67.7%;  Score 1048.6;  DB 2;  Length 175424;
Best Local Similarity  83.1%;  Pred. No. 2.1e-188;
Matches 1224;  Conservative   0;  Mismatches  234;  Indels  15;  Gaps   2;

Qy        77  GGTATCAGAATGGCAAAGCAAACTTAACCTTCAGAAAATACAGATGATTGGACAAAAGAG  136
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    170584  GGTATCAGATGGGCAGCACAACTTAACTTGCCAAAAAATACGGATGACTGGACCAAAGAG  170525

Qy       137  GATGTAAATCAGTGGTTAGAAAGTCATAAGAATGACCAAAAACACAGGGAAATTTGACT  196
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    170524  GATGTAAACCGGTGGTTAGAAAGTCATAAAATTGACCAAAAATTGACCAAAAACACAGGGACATTTGACT  170465

Qy       197  GAACAAGACGTGAATGAGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATG  256
              ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    170464  GCACAAGATGTGAATGGCGCCAAACTTAGATACTTAACTTAACTTAAGAACACCTTATTGATATA  170405
```

```
Qy    257 GGCATCACACATGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGAAAACA 316
          ||||| |||||||||||||||| | |||||||  |||||| |||||||| |||||||
Db 170404 GGTATAACATTTGGGCCAGCTATCCGATAGAACATTTATTCAAAGAGTTGATGGAAACA 170345

Qy    317 GCCATTGAAGATTCGATTCAGACATCTAGATGGGAAAGCCCAGTAAAAATGCTCCTAA 376
          |||||| |||||  | ||||||||  ||||||| |||||||  |||||||| ||||||
Db 170344 TCCACTGAAGATCCTTCTCAGATTGTTAAGGAGAAGCAAAAACAAAGGGTAAA       170285

Qy    377 GACCAAACTGTGTCTCAAAAGGAACGTAGAGAACTTCAAACAAAACAAAGGGTAAA    436
          |                 | ||||| |||||| ||||||||  || ||||||||||
Db 170284 AA---------ACAAGAGAAAAACAGAGAAACTTCAAAGCAAAACAAAAGGATAAA     170237

Qy    437 GAGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCA 496
          |||||||||||| |||| ||||| ||||||||  |  ||| || |||| ||  |||| |
Db 170236 GAGACATCAGATAGAGTTACTGACTCTACAATAAGTACAGTTACTCGAAGGTTCTATGTCA 170177

Qy    497 CTAAAAGTTGAGCTCATAGAAGATAAATATAGATTATACAAAGGAAAGGCAACCATCCATA 556
          |||| || |||||| ||| || || ||||| || | ||||||||||||  ||||| |
Db 170176 CTTAACAATGAATTCATGGAAGATGAAATAAATGACACAGAAAAAGCAGTCATCGTTA     170117

Qy    557 GACCTGACATGTGTATCATATCCATTGGATGTAATCCATTCCAGTAATCCATATCGTTACAAGTTG 616
          |||| |  |||||||  || ||||||  |||| ||||||||| |||||||||  |||||||||||
Db 170116 GAACTGATGTGTGTACCATACCCTTTTGATGTAATCCATTCAGTAATCCATATCGTTACAAATTG 170057

Qy    617 GATTTTAGTCTACAGACCTGAAACAGAGGACCAGGCAACATTCATTCATTGATCCGATACGAATTC 676
          |||| |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db 170056 AATTTTAGTCTACAGAGCTGAAACAGAGCAGCACCACTCAATCCAATACATGAATTC            169997

Qy    677 AAAGCCTTCACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTAGCAAT 736
          |||||||||||| |||||||||||||||| ||||||| |||||||||||||||||||
Db 169996 AAAGCCTTCACAAACAAACAGCAAACAGCAGCCACAAAGAGGATGTTAAGATGAAATTAGCAAT 169937

Qy    737 GAGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTT 796
          ||||| |||||||||||||||||||||||||||| ||||||||||||||||| |||||||
Db 169936 GAGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACCATTCATTTT 169877

Qy    797 GGAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAGTCACCAAATGATACCAAG 856
          ||||||||||  |||||||||||||||||||||  |||||||||||| |||| |||||
Db 169876 GGAGTCAAGGACAAACCCATGGGAAAATTGTTGGCCTGGAAGTCACCACTGTCACCAAG 169817

Qy    857 GAAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTGAAGACCATCCAAGTC 916
          ||||||| || || | |||||| ||||||| ||||||||  ||||||||||||| ||||
Db 169816 GAAGCCTTCATTCATCCACCACTCAATATGTGATCCACCAGTATTTGAAGACCATCAGGTC 169757

Qy    917 CAACAAGCAAAGAAGTGCATTCGATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGT 976
          ||||||||||||  ||||||||||||||||||||||| ||||||||||| ||||||||||||
Db 169756 CAAAAGCGAAGAACTGCATTCGAGAGCCAAGATTTGTAGAAGTTTACTGCCAAACAGT      169697

Qy    977 ACTCTATCTGACAGATTTGTTATTGGACATTATTCCACAGTTCTCTGAATGCCAA 1036
          ||| | | | |||| |||||||||| |||||||| ||||| | ||||||||||||
Db 169696 ATTCCATCTGACAGATTTGTTATTGGATGTTGTTCCAAAATATTCTGAATGTGAA     169637

Qy   1037 TATGATTACTTCCAGATTAAAATGCAAAATGCAAAATAAATGCAACAACAAAATATGGAACAAGTAAA 1096
          |||| |||| |||||||||||| |||||||| || | |||| |||||||| ||| ||||||||| ||
Db 169636 CAAGAGTATTCCAGATTAAAATGCAAAATTTAAAAATGCAAAACATCAGGACAACAAGACAATGGAAACAAAGTCCA 169577

Qy   1097 AAATTCTCACTATTTGTGTGGAGATGGACGTCTAAGGACCATTACGAAAAAATAAAGTT 1156
          ||||||||||| || | |||||||||||| ||||||||||||| | ||||||||||
Db 169576 AAATTCTCAGTCTTCATGCGAGATGGGGCCAGTTCAAGGACGACCAGTTGAAGCAAAAAATAATGTA 169517

Qy   1157 GATTTCAGAGCAGCATTTAAAGCAGATTTTAAAAACTGGCACTGGCGAGAGTCCAGAAAGCAGCAGAGA 1216
          ||||||||| |||||||||||||||||| ||||||||||| | || |||| |||||||| ||||| ||
Db 169516 GAATTCAAAACATTTAAGCTGATTTTAAAAATAAAAAAAAACTGCGAGCTAGGAAGAGAGGCCAGAA 169457

Qy   1217 AAATTCTCAGAGCAAAAACAAAACAAAATAAAAAAAAAGAAAAAGAGGGACCAAGTTGGTTAAATTA 1276
          ||||||||||||||||| |||| |||||| ||||||||| ||||||| |||||| |||||| ||
Db 169456 GAAAATGCGAGACTGACAGCAAAACAAACAAAAAAGAGAGTGAGGACCAAAAACTTGTTAAATTG 169397

Qy   1277 TTGACAGGAAATACAAGATTTGTTAGATATTCATATGAACAGTACATTCTTGTAACA 1336
          |||| ||||||| ||||| |||||||| || || || | || ||||  ||| |||
Db 169396 TTAACAGGAAATAAGGATTTGTTAGATAATTCATCATGACTATGACTATCTTGTAATA 169337
```

```
Qy   1337 AATAAATGCCACCCAGATCAAACAAACAAAACACTTAGATTCCTGAGGAAATTAAATGGTTT 1396
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||  ||||||
Db 169336 AATAAATGCCATCCAGCTGATAAAACTTAGATTCCTAAAGGAAATTAAGGGTTT         169277

Qy   1397 GCTGCTATTGGAGTTTGATCCTGAGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAA 1456
          | | || || ||  ||||||||||| ||| |||||||||||| |||||||||||||||||
Db 169276 GCTGTGTTGGAGTTTGATCCTGAATCTGAGTCGTGAGTGAAGGTGTGGTCAAAGCTTACAAAGAGA 169217

Qy   1457 AGCCGAGTAGCAAACCTCACTTCCAAGTGTATATGTGACAGAAACCACCACCAAAT 1516
          |||||||||||||||| | || |||||||  || ||| ||||  || ||||||||
Db 169216 AGCCGAGTAGTTAACCTTCCAAGTTTATATGTGGAGGGGAAACCA---CAAAT     169160

Qy   1517 GAGACGATTTCTACTCTAAAATCTTTACCATCAA 1549
          ||||| |||| ||  || |||||  || | ||||
Db 169159 GAGAAGATTATTAGTCTGAATCTTTTATCAACAA 169127
```

RESULT 10
AC108893/c

```
LOCUS       AC108893     108538 bp    DNA     linear   HTG 19-JUL-2002
DEFINITION  Bos taurus clone RP42-500G8, WORKING DRAFT SEQUENCE.
ACCESSION   AC108893
VERSION     AC108893.2  GI:21909465
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 108538)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maskeri,B., Mastrian,S.D.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 108538)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3 (bases 1 to 108538)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
            On Jul 19, 2002 this sequence version replaced gi:18464069.
COMMENT     -------------- Genome Center
              Center: NIH Intramural Sequencing Center
              Center code: NISC
              Web site: http://www.nisc.nih.gov
              Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
              Center project name: cjo
              Center clone name: 500G08

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
```

with a Phrap-derived quality score.
------- Summary Statistics
    Sequencing vector: Plasmid; n/a; 100% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.990319
    Consensus quality: 108439 bases at least Q40
    Consensus quality: 108499 bases at least Q30
    Consensus quality: 108531 bases at least Q20
    Insert size: 105000; agarose-fp
    Insert size: 108538; sum-of-contigs
    Quality coverage: 13.02x in Q20 bases; agarose-fp
    Quality coverage: 12.60x in Q20 bases; sum-of-contigs
----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*   1 108538: contig of 108538 bp in length.

FEATURES   Location/Qualifiers
     source   1..108538
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /clone="RP42-500G8"
                     /clone_lib="RP42"
     misc_feature   1..108538
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left
                     clone_end:T7
                     vector_side:right"

ORIGIN

Query Match   67.1%; Score 1039.8; DB 2; Length 108538;
Best Local Similarity 83.2%; Pred. No.1e-186;
Matches 1227; Conservative   0; Mismatches 227; Indels 21; Gaps   3;

Qy   75 TGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAG 134
            |||||||||||||||||||||||| |||||||||||||||||||||||| |||||||||||
Db   53771 TAGGTATCAGAATGGCAGCACAACTTAACTTACCAGAAAATACAGATGACTGACAAAAG 53712

Qy   135 AGGATGTAAATCAGTGGTTAGAAAGTCATAAGAGTTGGTTAAAAAGAACATCTTGTTGATA 194
            |||||||||||||||||||||||| ||||||| || |||||||||| || ||||||| ||
Db   53711 AGGATGTAAATCAGTGGTTAGAAAGTCATAAAATTGAACAAAACACAGAGAGCAATTTTGA 53652

Qy   195 CTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTAAAAAAAAGAACATCTTGTTGCTA 254
            |||||||||||||||||||| ||| ||||||||||||||||||||| ||||||||||||||
Db   53651 CTGCACAAGATGTGACTGGTGCAATCTTGAAGTACTTAACTACAGAGATCACCTTGTTGCTA 53592

Qy   255 TGGGCATCACCATGACCAGCTATTCAAATAGAAGAACTATTCAAAGGAATTGCGGAAA 314
            ||||||||||||||||||||| |||||||||||||| ||||||||||||||||||||
Db   53591 TGGGCATAACATTTGGACCAGCTATCCAAATAGAACAGCTTCTCAAAGAGTTGCTGGAGA 53532

Qy   315 CAGCCATTGAAGATTCGATTCAGACATCAAGATGGGAAAGCCCAGTAAAAATGTGCTCCTA 374
            |||||||||||||||||| ||||||| |||||||| |||||||||||||||| |||||||
Db   53531 CATCCTCTGGAGATCCTTTCCAGACATTGAAGAGCGAAAGGCAGTAAAAGTGTTCCTA 53472

Qy   375 AAGACCAAACTGTGTCTCAAAAGGAACGTAGAGCGTAAGAGAAAACTTCAAGCAAAAAGGGTA 434
            || |||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db   53471 AA--------ACACAGCAGAAGATGGAGGGACTTC-----AAAACAAAGAATA 53430

Qy   435 AAGAGAACCCAGAATGGCTAATCCGTCGAATGATGCTCGCAATGATTATCAAAGGAAAGGCAACCATCCA 494
            ||||||||||||| ||||||||||| |||||||||||| |||||||||||||||||||||||||||||||
Db   53429 AAAAGAAATCAGATAAGGTTAATGACTCACAGTGAGTACCTTACAGTAGGTTCTAAGT 53370

Qy   495 CACTAAAAGTTGAGCTCATAGAAGATAAAATAGATTATCAAAGGAAAGGCAACCATCCA 554
            |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   53369 CACTGAAAAATGAGTTCATCGAAGATGAAATAGATGAAATAGATGAAAAAGAGCAGCCATCCA 53310

Qy   555 TAGACCTGACATGTGTATCATATCCATTGATGAATTCAGTAATCCATATCGTTACAAGT 614
            ||||||||| || || |||||||||||||||||||| ||||||| ||||| ||||||||
Db   53309 CAGAACCGACACGCATGGCATACCCTTTGATGATGTTTGTGATGAATTCAGTGATCCATATCGTTACAAAT 53250

Qy   615 TGGATTTTAGTCTACAGCCTGAAACAGGACCAGCCAATCTCATTGATCGATACATGAAT 674
            ||||||||| |||||||||||||| ||||||||||||||||||||||| ||||||||||
Db   53249 TGAATTTAAACTACAGCCTGAAACAGGACCACTCAATCTCATTGATCCAATACATGAAT 53190

Qy   675 TCAAAGCCTTCACAAATACAGCAACGCCACAGAGGATGTCAAGATGAAATTAGCA 734
            || |||||||||||||||||| ||| ||||||||||||||||||||||||||||||
Db   53189 TCAAAGCCTTCACAAACACAGCAACAGGGCACGGAGGATGCTAAAATGAAATTAGCA 53130

Qy   735 ATGAGGTTTCCGATTTGCTTCAGCTTGTGATGAATTCACGTACCAATGGCACTATTCATT 794
            ||||||||||||||||||||||| |||||||||| ||||| ||||||||||||||||||
Db   53129 ATGAGGTGTTCCGATTTGCTTCTCTGTGTATGAACTCACGTACCAATGGCACCATTCATT 53070

Qy   795 TTGGAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAATGATACCA 854
            |||||||||||||||||||| |||||| |||||||||||||||||||||||||||||||
Db   53069 TTGGAGTCAAGACAAACCTCATGGAACAATTGTTGGCGTGGAATTTACCACCATCACCA 53010

Qy   855 AGGAAGCCCTCATTAACCATTCAATCTGATGATAAACAAGTATTTGAAGACCATCAAG 914
            ||||||||| ||||||||||| ||||||||||| |||||||||||||| |||||||||||
Db   53009 AGGAAGCCCTCATTGACCACTTCAATCTGATCCACCAATATTTTGAAGACCATCAGG 52950

Qy   915 TCCAACAGCAAAGAGAGTGCATTCGAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATA 974
            |||||||| ||||||| ||||||||||||||||||| |||||||||||||||| |||||
Db   52949 TCCAAAAAGCAAAGAGAACTGCATTCGAGAGCCCAAGATTTTACTGCGCAAACA 52890

Qy   975 GTACTCTATCTGACAGATTTGTTATTGGACATATTCCAGTTCCTGAATGCC 1034
            |||||||||||||||||||||||| |||||| |||||||||||| ||| ||
Db   52889 GTACTCTATCTGACAGATTTGTTATTGAAGTGGATGTTGTTCCAAATATTCTGAATGTG 52830

Qy   1035 AATATGATTACTTCCAGATAAGTGCAAAATGCAAAATCAAAAATAATGGAACAAAGTA 1094
            ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   52829 AAGTTGATTATTTCCAGATAAATGCAAAATGCAAAATGCAACAATGAAACAAAGTC 52770

Qy   1095 AAAAATTCCACTATTTGTGCGGATGGGACCAGCTTCTAAGGACATTACGAAAAATAAAG 1154
            |||||||||||||||||| ||||||||| |||||||||| |||||||||||||||||
Db   52769 CAAACTTCTCAGTCTTTTGTGCGGAGATGGGGCCAGTTCTAACGACATCATGAAAAAGTAATA 52710

Qy   1155 TTGATTTCAGAGCATTTAAAGCAGATTTAAAACACTGGCGAGTCCAGAAAAGCAGCAG 1214
            |||||||| ||||||||||||||||| ||||| ||||||||||||||||||||||||||
Db   52709 TGGATTTCAAAGCATTTAAATTTAAATTAGATTTGAAAAGTCACACTTCAGGAAAGGGCAG 52650

Qy   1215 AAGAAAATTCGAGACAAAGACAAATAAAAAAAGAAGAGAGGGACCAAAGTTGGTTAAAT 1274
            |||||||| |||||||||||||| ||||| ||||||||||| ||| ||||||||||||||
Db   52649 AAGAGAAAATGCAGAGTGAAAACAAAATAAAAGAAGCGAGGGGCCAAAGCTTGTTAAAT 52590

Qy   1275 TATTGACAGGAAATCAAGATTTGTTAGATAATTCATCATATGAACAGTACATTCTTGTAA 1334
            |||||||||||||||| ||||||||| |||||| ||| ||||| ||||||||||||||||
Db   52589 TGTTGACAGGAAATCAGGATTTGTTTAGATAATTCATCATACAACTGGTACATTCTTGTAA 52530

Qy   1335 CAAATAAATGCCACCCAGATCAAATAAAGAGGAATTAGATTTCCTGAGGAATTAAATGGT 1394
            ||||||||||||||||||| |||||||| |||||||| |||||||||||||||||||||||
Db   52529 TAAATAAATGCCACCCTACACAAATCAAACTTAGATTTCCTAAAGGAATTAAATGGT 52470

Qy   1395 TTGCTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGAGTCTAAAGTTGGTCAAAGGT 1454
            ||||||| ||||| ||||||||||| ||||| |||||||| ||||| ||||||||||
Db   52469 TTGCTGTGTGTTGGAGTTTGATCTGAATCTGTGTGAGCAAGGGTGTGTCGTCAAAGCATACAAAG 52410

Qy   1455 AAAGCCGAGTAGCAAACCTTCACTTTCCAGTGTATATGGAAACAAAACACACAA 1514
            ||||||||||||||||||||||| |||||| |||||||||||| ||||||||||
Db   52409 AAACCCGAGTAGTAAATCTTCACTTCCCAAGTTGTATGTGGAGGGGGAAAACCA----CAA 52353

Qy   1515 ATGAGACGATTTCTACTCTAAATCTTTACCATCAA 1549
            |||||||||||||||||||||| ||||||||||||
Db   52352 CTGAGAAGGATTACTAGTCTGAATCTTTATCACAA 52318

RESULT 11
AC079885/c
LOCUS   AC079885   148278 bp   DNA   linear   HTG 11-SEP-2002

```
DEFINITION   Rattus norvegicus chromosome 4 clone RP31-246H18 strain Brown
             Norway, WORKING DRAFT SEQUENCE, 6 ordered pieces.
ACCESSION    AC079885
VERSION      AC079885.3  GI:22779511
KEYWORDS     HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE       Rattus norvegicus (Norway rat)
ORGANISM     Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 148278)
AUTHORS      Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
             Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
             Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
             Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
             Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
             Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
             Mccloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
             Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schueler,M.G., Sison,C.,
             Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
             Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
TITLE        NISC Comparative Sequencing Initiative
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 148278)
AUTHORS      Green,E.D.
TITLE        Direct Submission
JOURNAL      Submitted (16-SEP-2000) NIH Intramural Sequencing Center, 8717
             Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE    3  (bases 1 to 148278)
AUTHORS      Green,E.D.
TITLE        Direct Submission
JOURNAL      Submitted (11-SEP-2002) NIH Intramural Sequencing Center, 8717
             Grovemont Circle, Gaithersburg, MD 20877, USA
             On Sep 11, 2002 this sequence version replaced gi:11120768.
COMMENT      -------------- Genome Center
               Center: NIH Intramural Sequencing Center
               Center code: NISC
               Web site: http://www.nisc.nih.gov
               Contact: nisc_zoo@nhgri.nih.gov
             -------------- Project Information
               Center project name: rk
               Center clone name: 246H18
             --------------
             The sequence data in this record represents an 'enhanced'
             version of a Phase 2 submission. Specifically, the indicated
             order and orientation of each sequence contig has been
             established using one or more of the following: read-pair
             data from individual subclones, overlaps with neighboring
             clones, alignment with available reference sequence (e.g.,
             human), and/or confirmation by PCR testing. In addition,
             the sequence assembly is based on at least 8X average
             coverage in Q20 bases and has been reviewed to rule out
             gross misassemblies, the low-quality ends of sequence
             contigs have been trimmed away, and each base is associated
             with a Phrap-derived quality score.
             -------------- Summary Statistics
               Sequencing vector: plasmid; n/a; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.990319
               Consensus quality: 146350 bases at least Q40
               Consensus quality: 147034 bases at least Q30
               Consensus quality: 147462 bases at least Q20
               Insert size: 145000; agarose-fp
               Insert size: 147778; sum-of-contigs
               Quality coverage: 11.71x in Q20 bases; agarose-fp
               Quality coverage: 11.49x in Q20 bases; sum-of-contigs
             --------------
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 6 contigs. Gaps between the contigs
             * are represented as runs of N. The order of the pieces
             * is believed to be correct as given, however the sizes
             * of the gaps between them are based on estimates that have
             * provided by the submittor.
```

```
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.
             *       1     18366: contig of 18266 bp in length
             *   18267     18366: gap of unknown length
             *   18367     42762: contig of 24396 bp in length
             *   42763     42862: gap of unknown length
             *   42863     83567: contig of 40705 bp in length
             *   83568     83667: gap of unknown length
             *   83668    121422: contig of 37755 bp in length
             *  121423    121522: gap of unknown length
             *  121523    127993: contig of 6471 bp in length
             *  127994    128093: gap of unknown length
             *  128094    148278: contig of 20185 bp in length.
FEATURES             Location/Qualifiers
     source          1..148278
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="Brown Norway"
                     /db_xref="taxon:10116"
                     /chromosome="4"
                     /clone="RP31-246H18"
                     /clone_lib="RP31"
     misc_feature    1..18266
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1..3899
                     /note="clone overlaps with GenBank Accession Number
                     AC079437 clone RP31-275K4 (center project name rj)"
     misc_feature    18367..42762
                     /note="assembly_fragment"
     misc_feature    42863..83567
                     /note="assembly_fragment"
     misc_feature    78240..148278
                     /note="clone overlaps with GenBank Accession Number
                     AC084240 clone RP31-103L21 (center project name rl)"
     misc_feature    83668..121422
                     /note="assembly_fragment"
     misc_feature    121523..127993
                     /note="assembly_fragment"
     misc_feature    128094..148278
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN
     Query Match       59.5%;  Score 921;  DB 2;  Length 148278;
     Best Local Similarity  77.6%;  Pred. No. 2.6e-164;
     Matches 1173;  Conservative  0;  Mismatches 290;  Indels  48;  Gaps   3;

Qy      72 TACTGGGTATCAGATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAA 131
           |||||| |||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db  114323 TACAGCTAATCAGATGGCAGAGACACTCAACCTTCAACAAAACACAGATGACTGGACAA 114264

Qy     132 AAGAGGATGTAAATCAGTGGTTAGAAAGTCATAAGTTCATAGATGATTGACCAAAAACACAGGGAAATT 191
           |||||||||| |||||| ||| |||||||| |||||||||||||||||||||||||||| |||| |||||
Db  114263 AAGAGGATGTAAACCTGTGGTTAGAAAGTCATAGATGATTGACCATAGATGATTGACCAAAAAGCACAGG 114204

Qy     192 TGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTG 251
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  114203 TGATGGCACAAGATGTGAGTGGAGTAGTCTTGAAGTGGTTAATTAAAAAAATCTTGTTG 114144

Qy     252 ATATGGCATCACACATGGACCAGCTATTCAAATAGAGAACTATTCAAAGAATTGCGGA 311
           ||||||||||||||||||||||||||| |||||||||||||||| ||||||||| ||||
Db  114143 AAATGGGCATAACCCATGGCCAGCTATCCAAATAGAAGAGCTATTCAAAGAATTGCAGA 114084

Qy     312 AAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGCGGAAAGCCCAGTAAAAATGCTC 371
           |||||||||||||||||||||||| |||||||| ||||| |||||||| ||||| |||||
Db  114083 AAACACCCAAAGAAAATCTTACTAAGACATGTAAGAACATGAAGAACAAAAGGCAGAAAAAATATCC 114024

Qy     372 CTAAAGACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAAAGCAAAAAG 431
           ||||| ||| ||| ||||
```

```
Qy      1479 TTCCAAGTGTATATGTAGAACAGAAAACCACACCACCAAATGAGACGATTTCTACTCTAAATC 1538
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    112898 TTCCAAGCCTATTTATAGAAGAGAAGAAAAAACCCACAGTAAGTGAAAAGATTTCCAGTCTGAATC 112839

Qy      1539 TTTACCATCAA 1549
             |||||||||||
Db    112838 TTTATCACGAA 112828
```

RESULT 12
AC084240/c
LOCUS   AC084240   180665 bp   DNA   linear   HTG 06-DEC-2002
DEFINITION Rattus norvegicus chromosome 4 clone RP31-103L21 strain Brown
            Norway, WORKING DRAFT SEQUENCE, 4 ordered pieces.
ACCESSION   AC084240
VERSION   AC084240.2 GI:26080471
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 180665)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paquirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE   NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 180665)
  AUTHORS   Green,E.D.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-OCT-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3 (bases 1 to 180665)
  AUTHORS   Green,E.D.
  TITLE   Direct Submission
  JOURNAL   Submitted (06-DEC-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT   On Dec 6, 2002 this sequence version replaced gi:10864175.
            -------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                Center project name: rl
                Center clone name: 103L21
            -------------- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 179812 bases at least Q40

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlap with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassembles, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
```

```
Db    114023 CTAAAACCCAGAC-----------AGAAAGTAGAGAGACTTCAAATCAAAACCAGAGAG 113976
                                        ||||||||||||||||||||||||||||||||
Qy       432 GTAAAGAGAACCCAGATATGCTAATCCGTCTGCAATG---------------------- 469
             |||||||||||||||||||||||||||||||||||||
Db    113975 CTGAAGAGAGACTCATGCAAGCCTGACACCTGCAGTGAGCTCCAGAGGAAACTCACAA 113916

Qy       470 --------AGTACAACTGCTAAAGGTTCTAAGTCACTAAAAGTTGAGCTCATAGAAG 518
                     |||||||||||||||||||||||||||||||||||||||||||||||
Db    113915 GGACTCCAGCAAGTCCAGAGCTAAAGGTTCTAAATCAGAAAAAAAATGACCTAGTAGAG 113856

Qy       519 ATAAAAATAGATTATACAAAGGAAAAGGCAACCATCCATAGACCTGACATGTGTATCATATC 578
             |||||||||||||||||||||||||                |||||||||||||||||||
Db    113855 ATAAAGCAGTCCA---ACAGGATAAGCAGGCATCTCCAGACCACCGTGCATTGCATATC 113799

Qy       579 CATTTGATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTTAGTCACAGCCTGAAA 638
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113798 CTTTCAATGAGTTCAGTAATCCATATCGTTACAAATTACATTTCATCCTTACAGCCTGAAA 113739

Qy       639 CAGGACCAGGCAATCTCATTGATCCGATACATGAATTCAAAGCCTTCACAAATACAGCAA 698
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113738 CAGGGCCACTCAATCTCATTGACCCAGTACAGTACGATGAGTTCAAAGCCTTAACACACAGAGA 113679

Qy       699 CAGCCACAGAGAGGATGTCAAGATGAAATTAGCAATGAAATTAGCAACGAAGTTTCCGATTTGCTTCAG 758
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113678 CAGCCACAGAGAGGAGACATAAAGATGAAATTAGCAATGAAATTAGCAACGAAGTCTTCCGATTTGCTTCAG 113619

Qy       759 CTTGTATGAATTCAGTACCAATGGCACTATTCATTTGGAGTCAAAGACAAACCCCATG 818
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113618 CTTGTATGAACTCACGTACTACTAATGGCACCATTCATTTGGAGTCCAAGACAAACCCCAG 113559

Qy       819 GGAAAATTGTTGGCATCAAAGTCACCAAGGAGCCCTCAATTGACCAATTCAACCATTCA 878
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113558 GAAGGATTGTTGGCATGGAGCTCACCACCGTCACTAAAGATGCCCTCATTGACCACTTCA 113499

Qy       879 ATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCAACAAGCAAAGAAGTGCATTC 938
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113498 ATCTGATGATAGTCCAGTATTTTGAGGTCCACCAGTCCACAACAAGCAAAGAAGTGCATTC 113439

Qy       939 GAGAGCCAAGATTGTGTTGGGAAGTTTTACTCGTCCAAATAGTACTCTATCTGACAGATTTGTTA 998
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113438 GAGAGCCAAGATTTGTGGAAGTTTTACTCGTCCAAATAGTACTCCATCTAATAAATTTGTTA 113379

Qy       999 TTGAAGTGGACATTATTCCACAGTTCTCTGAATGCCAATATGATTACTTCCAGATTAAAA 1058
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113378 TTGAGGTGGATATCATTCCACAAATATCCACAATATTCAGAATGCAAAGAAGAGATTTCTTCCAAATAAAAA 113319

Qy      1059 TGCAAATTACAACAACAAATATGGAACAAAGTAAAAAATTCCACTATTTGTGCCAG 1118
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113318 TGCATTGGCATAAGGATGAAACATGCCAACAAATTCCAAATTACTCGGTCTTTCTGAGAG 113259

Qy      1119 ATGGGACCAGCTCTAAGGACTTACGAAAAATAAAGTTGATTTCAGAGCATTTAAAGCAG 1178
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113258 ATGGTCCGAAAGTAAGGAAGCAACATCATAGGGAATAAAGCTGATTTCAAAGCTTTAAGTTAG 113199

Qy      1179 ATTTTAAAACACTGGCAGAGTCCGAGAAAAGCAGCAGAAGAAGAAAAAAATTCAGAGCAAAAAACAA 1238
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113198 ACCTAAAAGAACTACTGGCAGACTCTAGGAAAGAAAGAGCAGAAAAAAAATCAAAGTAAAATCAG 113139

Qy      1239 ATAAAAAAGAAGAGAAGAGAGGGACCAAAGTTGGTTAAATTATTGACAGAAATCAAGATTTGT 1298
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113138 ATGGAAGCAACAGTCAGGGGGCAAAAGCTAATTGACCTGTTGACAGGAAATAAGGATCTGC 113079

Qy      1299 TAGATAATTCATACTATGAACAGTACATTCTTGTAACAAATAAATGCACCCAGATCAAA 1358
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113078 TAGATAACTCACACTATGGTACATTCTTGTAACAAATAAATGTCATCCAGATCAAA 113019

Qy      1359 CAAAACACTTAGATTTCCTGAAGGAAATAAATGTTTGCTGTATTGGAGTTTGATCCTG 1418
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113018 CCAAACACCTAGATTTCCTGAAGGAAATTAAATGTTTGCTGTACTAGAGTTTGATCCTG 112959

Qy      1419 AGTCTACATCAATGAGTGGTCAAGCTTCAGAGAAAGAGCCGAGTAGCAAACCTTCACT 1478
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    112958 AGTCTGAGAGCAATGGTGTGTGGTCAAGGAGCTTTCAAGGAAGCGAGTAGCAAACCTTCACT 112899
```

```
Db  37195  AAACACCCAAAGAGAAATCTTACTAAGACATGTAAGAGACATAAAGGCAGAAAAAAATATCC  37136
                  |||||| ||||    | ||||  |||||||||||||||||||||||||  |||||||||
Qy    372  CTAAAGACCAAACTCGTCTCAAAGAACTTCAAAGAACTAGAGAAAACTCAAAGCAAAACAAAAG   431
           |||||| |||||
Db  37135  CTAAAACCCAGAC------------AGAAAGTAGAGACTTCAAATCAAACCACGAGAGAG  37088
                  |||||| |||||   |||||||||||||||||
Qy    432  GTAAAGAGAACCCAGATATGGCTAATCCGTCTGCAATG---------------------   469
           |||||||||||||||||||||||||||||||||||
Db  37087  CTGAAGAGAACTCATGCAGCTGACACCTCTGCAGTGAGTCCAGAAGGAAACTCACAAA  37028
                  ||||||||||||||||||||||||||||||||||||||||||||
Qy    470  ---------AGTACAACTGTAAAGGTTCTAAGTCACTAAGTTGAGTCATAGAAG   518
                       |||||| |||| ||||||||||||||||
Db  37027  GGACTGCGAAGTCCGAGACGTCAAAGGTTCTAAATCAGAAAAAAATGACCTAGTAGAAG  36968
                 ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Qy    519  ATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTGACCATGTGTATCATATC   578
                   |||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db  36967  ATAAAGCAGTCCA---ACAGGATAAGCAGGCATCTCCAGACCAACCAACCTGCATTGCATATC  36911
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    579  CATTTGATGAATTCAGTAATCCATATGTTACAGAGTTGGATTTTAGTCACAGCCTGAAA   638
                 |||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db  36910  CTTTCAATGAGTTCAGTAATCCATATCGTTACAAATTACATTTCATCCTACAGCCTGAAA  36851
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    639  CAGGACCAGGCAATCTCATTGATCCGATACATGATTCAAAGCCTTCACAAAATACAGCAA   698
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36850  CAGGGGCCACTCCAATCTCATTGACCCAGTGACCACATGAGTTCAAAGCCTTAACAAACACAGAGA  36791
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    699  CAGCCACAGAGACGAGGATGTCAAGATGAATGTCAAGATGAAATTTAGCAATGAGGTTTCCGATTTGCTTCAG   758
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36790  CAGCCACAGAGACGAGGACCATAAAGATGAAATTTAGCAACGAAGTCTTCCGATTTGCTTCAG  36731
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    759  CTTGTATGAATTCACGTACCAATGGCACTATCATTTTGGAGTCAAAGACAAACCCCATG   818
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36730  CTTGTATGAAGTCACGTACTACTATCATTTTGGAGTCCAAGAGCAAGACAAACCCCAAG  36671
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    819  GGAAAATTGTTGTGGCATCAAGTCACCAATGATACCAAGGAAGCCCTCATTAACCATTTCA   878
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36670  GAAGGATTGTTGTGGCATGGAGTCACCACCGTCACTAAAGATGCCCTCATTGACCACTTCA  36611
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    879  ATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCAACAGCAAGAAGTGCATTC   938
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36610  ATCTGATGAGTCCAGTATTTTGAGGTCCATCAAGTCCAACAGCAAGAAAGGAAGTGCATTC  36551
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    939  GAGAGCCAAGATTTGTTGGAAGTTTTACTGCCAAATAGTACTCTATCTGACAGATTTGTTA   998
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36550  GAGAGCCAAGATTTGTTGGAAGTTTTACTGCCAAATAGTACTCCATTAATAAAATTTGTTA  36491
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    999  TTGAAGTGGACATTATTCCACAGTTCTCTGAAGATGCCAATATGATTACTTCAGTTAAAA  1058
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36490  TTGAGGTGGATATCATTCCACAATATTCAGAATGCAAGAAGATTTCTTCCAAATAAAAA  36431
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1059  TGCAAAATTACAACAACAACAAAATATGGGAACAAAGTAAAAATTCCTCACTATTGTGTGCGCAG  1118
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36430  TGCATTGGCATTAAGGATGAAACATGGCAACAAATTCCAAAATACTCCGGTCCTTGTGTGAGAG  36371
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1119  ATGGGACCAGCTCTAGGACATTACGAAGAAATAACAGATTGATTTCAGAGCATTTAAAGCAG  1178
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36370  ATGGTCCGAAAGTAGGACATCATAGGGATTAAAGCTGAAGAAAATTCAGAGCAAAAACAG  36311
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1179  ATTTTAAAACACTGGCAGAGTCCAGAAAGCAGCAGAAGAAAAATTCAGAGTAAAATCAG  1238
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36310  ACCTAAAAGCACTGGCAGACTCTAGGAAGAAAGCACAGAGGCAGGGAAATCT  36251
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1239  ATAAAAAAAGAAAGAAAGAGGGGACCAAAGTTGGTTAAATTATTGACAGAGATTTGT  1298
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36250  ATGGAAGCAACAGTCAGGGGCAAAAGCTAATTGACCTGTTGACAGAAATAGGATCTGC  36191
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1299  TAGATAATTCATCTATGACAGTACATTCTTGTAACAAATAAATAAATGCCACCAGATCAAA  1358
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36190  TAGATAACTCACCATGAATGGTACATTCTTGTAACAAATAAATAAATGTCATCCAGATCAAA  36131
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1359  CAAAACAACTTAGATTTCCTGAAGGAAATTAAATGGTTTGCTTATTGGAGTTGATCCTG  1418
                 ||| ||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  36130  CCAAACAACTTAGATTTCCTGAAGGAAATTAAATGGTTTGCTTATTGGAGTTGATCCTG  36071
```

FEATURES   Location/Qualifiers
source   1..180665
                  /organism="Rattus norvegicus"
                  /mol_type="genomic DNA"
                  /strain="Brown Norway"
                  /db_xref="taxon:10116"
                  /chromosome="4"
                  /clone="RP31-103L21"
                  /clone_lib="RP31"
misc_feature   1..71303
                  /note="clone overlaps with GenBank Accession Number
                  AC079985 clone RP31-246H18 (center project name rk)"
misc_feature   1..5391
                  /note="assembly_fragment
                  clone_end:SP6
                  vector_side:left"
misc_feature   5492..51132
                  /note="assembly_fragment"
misc_feature   51233..146593
                  /note="assembly_fragment"
misc_feature   146694..180665
                  /note="assembly_fragment
                  clone_end:T7
                  vector_side:right"

ORIGIN

Query Match   59.5%;   Score 921;   DB 2;   Length 180665;
Best Local Similarity 77.6%;   Pred. No. 2.5e-164;
Matches 1173;   Conservative   0;   Mismatches 290;   Indels 48;   Gaps   3;

```
Qy     72  TACTGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAAATACAGATGATTGGACAA   131
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37435  TACAGCTAATCAGAATGGCAGAGACACTTCAACCTTCCAAAAAACCACAGATGACTGGACAA  37376
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    132  AAGAGGATGTAAATCAGTGGTTAGAAAGTCATAAAGTTCATAAAGTTGAAAGTTGGGAAATT   191
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37375  AAGAGGATGTAAACCTGTGTGGCTAGAAAGTCATAAAGTTCATAAAGCACAGAGGCAGGGAAATCT  37316
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    192  TGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAAAGAAACATCTTGTTG   251
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37315  TGATCGCACAAGATCGTGAGTGGTGAGTAGTCTTGAAGTGGTTAAATAAAAAAATCTTGTTG  37256
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    252  ATATGGGCCATCACACATGGCCAGCTATTCAAATAGGAAGAAACTATTCAAAGAATTGCGAA   311
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  37255  AAATGGGCATAACCCATGGCCAGCTATCCAAATAGCAGAGAGCTATTCAAAGAATTGCAGA  37196
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    312  AAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAGCCCAGTAAAAATGCTC   371
                 |||||||  ||  |  |||||||||   ||    ||||||| |  |||
Db  36130
```

NOTE: This is a 'working draft' sequence. It currently
consists of 4 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
provided by the submittor.
This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.
      1   5391: contig of 5391 bp in length
   5392   5491: gap of unknown length
   5492   51132: contig of 45641 bp in length
  51133   51232: gap of unknown length
  51233   146593: contig of 95361 bp in length
 146594   146693: gap of unknown length
 146694   180665: contig of 33972 bp in length.

Consensus quality: 180152 bases at least Q30
Consensus quality: 180275 bases at least Q20
Insert size: 138000; agarose-fp
Insert size: 160000; pulse-field-gel
Insert size: 180365; sum-of-contigs
Quality coverage: 15.11x in Q20 bases; agarose-fp
Quality coverage: 13.03x in Q20 bases; pulse-field-gel
Quality coverage: 11.56x in Q20 bases; sum-of-contigs

```
Qy   1419 AGTCTAACATCAATGAGTGGTCAAAGCTTACAAAGAAAGCCGAGTAGCAAACTTCACT 1478
          ||||||  |||||||| |||| |||||  ||| ||||| |||||||||||||| | ||||
Db  36070 AGTCTGAGAGCAATGTGTGTCGGTCAAGGCTTTCAAAGGAGCCGAGTAGCAAACCTCCACT 36011

Qy   1479 TCCAAGTGTATATGTAGAACAGAAACCACACCAAATGAGACGATTTCTACTCTAAATC 1538
          |||||| ||||  |||| |||| |||| |||| | ||||||  |||||||  | |||||
Db  36010 TCCAAGCCCTATTTATAGAGAAAAAACCCAGTAAGTGAAAAGATTTCCAGTCTGAATC 35951

Qy   1539 TTTACCATCAA 1549
          ||| |||  ||
Db  35950 TTTATCAGCAA 35940
```

RESULT 13
AC134055/c

LOCUS   AC134055   221557 bp   DNA   linear   HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-158021, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC134055
VERSION   AC134055.3 GI:30522024
KEYWORDS   HTG; HTGS_PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 221557)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,K., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Catter,K., Cavazos,I., Ceasar,H., Center,A., Chen,J.,
            Chacko,J., Chavez,D., Chen,G., Chen,H., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,C., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado.R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,R., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rockey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,A., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,H., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wieczyk,R., Wooden,H., Worley,K.,
```

```
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 221557)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 221557)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24955744.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the contig and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            --------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            --------------- Project Information
            Center project name: GHAT
            Center clone name: CH230-158021
            --------------- Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 215241 bases at least Q40
            Consensus quality: 217114 bases at least Q30
            Consensus quality: 218453 bases at least Q20
            Estimated insert size: 248512; sum-of-contigs estimation
            Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            ---------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1         5505: contig of 5505 bp in length
            *       5506         5605: gap of unknown length
            *       5606       220233: contig of 214628 bp in length
            *     220234       220333: gap of unknown length
            *     220334       221557: contig of 1224 bp in length.
FEATURES             Location/Qualifiers
     source          1..221557
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-158021"
     misc_feature    1..2522
                     /note="wgs_end_extension
                     clone_end:F7"
     misc_feature    232..1034
                     /note="clone_boundary
```

```
                    clone_end:T7
                    site:EcoRI
misc_feature        end_sequence:BH313763"
                    2573. .5505
                    /note="wgs_contig"
misc_feature        36039. .62916
                    /note="clone_boundary
                    clone_end:Sp6"
                    site:EcoRI
                    end_sequence:BH313765"
misc_feature        218385. .220233
                    /note="wgs_end_extension
                    clone_end:Sp6"

ORIGIN
Query Match         59.5%;   Score 921;   DB 2;   Length 221557;
Best Local Similarity   77.6%;   Pred. No. 2.4e-164;
Matches 1173;   Conservative   0;   Mismatches 290;   Indels 48;   Gaps   3;

Qy      72 TACTGGGTATCAGAATGGCAAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAA 131
           || || | ||||||||||||  || |||||  || |||| ||||||| ||||||||||||
Db  100935 TACAGCTAATCAGAATGGCAGAGACACTCAACCTTCCAAAAAACACAGATGACTGGACAA 100876

Qy     132 AAGAGGATGTAAATCAGTGGTTAGAAAGTCATAGAAGTCATAGAAACACAGGGAAATT 191
           ||||||||||||  || |||||||||||  |||| ||||| ||||||||||||||
Db  100875 AAGAGGATGTAAACCTGTGGCTAGAAAGTCATAGGAGTCATAGAAGCACAGGGAAATCT 100816

Qy     192 TGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTG 251
           ||  ||||||||| |||||||| |||||||||||||||| |||||| ||||||| |||
Db  100815 TGATGGCACAAGATGTGAGTGGAGTAGTCTTGAAGTGGTTAAATAAAAAAATCTTGTTG 100756

Qy     252 ATATGGGCATCACCATGGACCAGCTATTCAAATAGAAGACTATTCAAAGAAATTGCGGA 311
           | |||||||||| ||||||| |||||||| |||| ||||| ||||||| ||||||
Db  100755 AAATGGGCATAACCCATGGGCCAGCTATCAAATAGAAGAGCTATTCAAAGAAATTGCAGA 100696

Qy     312 AAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAATGCTC 371
           ||||| ||  |||||||| ||  ||||||||||||| ||||| |||||||||||||| |
Db  100695 AAACACCCAAAGAAGAAATCTTACTAAGACATGTAAGATGTAAGAGCCCAGTAAAATATCC 100636

Qy     372 CTAAAGACCAAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAAAACAAAAGG 431
           || |||                ||||||||| ||||||||||||||| ||||| ||||||
Db  100635 AAACACCCAAAGAAATCTTACTAAGACAATCTTACTAAGGAGACAAAAGCAAAAACAAATCC 100588

Qy     432 GTAAAGAGAACCCAGATATGGCTAATCCGTCTGCAATG-------- 469
           |||||              || ||||| ||||||||||
Db  100587 CTGAAGAAACTCATGCATGCGCAAGGCTGACACCCTCTGCAGTCGAGTCCAGAAGGAAACTCACAA 100528

Qy     470 ----------AGTACAACTGTCTAAAGGTTCTAAGTCACTAAAAGTTGAGCTCATAGAAG 518
                     ||||||||||||||||| ||||| ||||||||| ||||||||
Db  100527 GGACTGCAGCAAGCTGCCAGAGGCTAAAGGTTCTAAATCAGAAAAAAAATGACCTAGTAGAAG 100468

Qy     519 ATAAAATAGATTATACAGGAGAAAGGCAACATCCATCAATAGACCTGACATGTGTATCATATC 578
           ||||||||||||||||||| |||||||||||| || ||||||||  ||
Db  100467 ATAAAGCAGTCCA--ACAGGATAAGCAGGCATCTCCAGGACCAACGTGCATTGCATTGCATATC 100411

Qy     579 CATTGATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTAGTCTACAGCCTGAAA 638
           |||| ||||||| |||||||||||||||| ||||||  ||| ||
Db  100410 CTTTCAATGAGTTCAGTAATCCATATCGTTACAAATTACATTTCATCTCCAGCCTGAAA 100351

Qy     639 CAGGACCAGGCAATCTCATTGATCCGATACATGAATTCAAAGCCTTCACAAATACAGCAA 698
           |||||||||||||| |||||| ||||| ||||||||||||||||||| |||||||||||
Db  100350 CAGGGCCACTCAATCTCATTGACCCCAGTACATGAGTTCAAAGCCTTAACAACACAGAGA 100291

Qy     699 CAGCCACAGAAGAGGATGTCAAGAGGCTCAAGATGAAATTAGCAATGAAGTTTCCGATTCAG 758
           ||||||| || || |||||||||||| |||||||||||||||||||| ||||||||||||
Db  100290 CAGCCGACAGAGGAGGACATAAAGATGCATAAGAAATTTAGCAACGAAGCTTCCGATTCCTCCAG 100231

Qy     759 CTTGTATGAATTCACGTACCAATGGCACTATTCATTTGGAGTCAAAGACAAAACCCCATG 818
           ||| ||| |||| ||||||  ||||||||||||||||| ||| ||||||| |||||||
Db  100230 CTTGTATGAACTCACGTACTAAATGGCACCATTCATTTGGGAGTCCAAGACAACACCCCAAG 100171

Qy     819 GGAAAATTGTTGGCATCAAAGTCACCAAGTACCACCACATGATACCAAGGAGGCCCTCATTACC 878
           ||     |    || ||| ||||| |||| |||| ||
Db        (continues)
```

```
Db  100170 GAAGGATTGTTGGCATGGAGCTCACCACGTCACCACGGTCACTAAAGATGCCCTCATTGACCACTTCA 100111
           || ||| ||||||||| |||||   || ||||||||| |||| |||||||| || |||||| |||
Qy     879 ATCTGATGATAAACAAGATATTTGAAGACCATCAGTCCAACAAGCAAGAGAAGTGCATTC 938
           |||| || |||| ||| |||||||||||||||| |||| |||||||||||||||||||||
Db  100110 ATCTGATGATAGTCCAGTATTTTGAGGTCATCAACAGTCCAACAAGCAAGAAGAAGTGCATTC 100051

Qy     939 GAGAGCCAAGATTTGTGGAAGTTTACTGCCAAATAGTACTCATCTGACAGATTTGTTA 998
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||| |
Db  100050 GAGAGCCAAGATTTGTGGAAGTTTTACTGCCAAATAGTACTCCATCTAATAAATTTGTTA 99991

Qy     999 TTGAAGTGGACATTATTCCACAGTTCCTGAATGCCAATGATGCCAATATTACTTCCAGATTAAAA 1058
           |||||||||||||| ||||| |||||||||  | |||| ||||||| |||||| ||||||||||
Db   99990 TTGAGGTGGATATCATTCCACAATTCCAGAATGCAATGCAAGAAGCAAAGAAGAGATTTCTTCCAAATAAAA 99931

Qy    1059 TGCAAAATTACAACACAAACAAATATGGGACAAAGTAAAAAATTCTCACTATTTGTGCGGAG 1118
           |||||||| || | || ||  || |||| |||  |||||| |   |||| || ||| |||||
Db   99930 TGCATTGGCATAAGGATGAAACATGGCAACAAATTCCAAAATACTCGGTCTTTGTGAGAG 99871

Qy    1119 ATGGGACCAGCTCTAAGGACATTACGAAAATAAAGTTGATTTCAGAGCATTTAAAGCAG 1178
           |||| || || |||||||| |||| |||||||||| | | ||||||| ||  ||||||||
Db   99870 ATGGTCCGAAAGTAAGGACATCATAGGGAATAAAGCTGATTTCAAAGCTTTTAAGTTAG 99811

Qy    1179 ATTTTAAAACACTGGCAGAGTCCAGAAAACAGCAGAGAAAAATTCAGAGACAAAAACA 1238
           | ||| |||||||||||||||  ||  ||||||||||||||||||||||||  |||||
Db   99810 ACCTAAAGCACTGCGCAGACTCTAGGAAAGAAGCAGCAAAGAAAAATCCAAAGTAAAATCAG 99751

Qy    1239 ATAAAAAAGAAAGAGGAGGACCAAAGTTGGTTAAATTATTGACGAGGAAATCAAGATTTGT 1298
           |||| || |||||||||||||||||||| ||| | ||||||  || || ||||| ||||
Db   99750 ATGGAAGCAACAGTCAGGGGCAAAAGCTAATTGACCTGTTGACAGGAAATAAGGATCTGC 99691

Qy    1299 TAGATAATTCATACTATGAACAGTACATTCTTGTAACAATAAAATGCCACCCAGATCAA 1358
           ||||||| |||| |||||||| |||||||||||||||  |||| ||| ||||||||||||
Db   99690 TAGATAACTCACACTATGAATGGTACAATCTTGTAACAATAAATGTCATCCAGATCAA 99631

Qy    1359 CAAAACACTTAGATTTCCTGAAGGAAATTAAATGGTTTTCGTGTATTGGTTGACTAGAGTTTGATCCTG 1418
           ||||||||| |||||||||||||| ||||  || |||||| || ||| | |||||||
Db   99630 CCAAACACCTAGATTTCCTGAAGGAAATTAAATGGTTTTCGTGTATTACTAGAGTTTGATCCTG 99571

Qy    1419 AGTCTAACATCAATGGAGTGGTCAAAGTTACAAAGAAGGTAGCAAACTTCACT 1478
           ||||| |||||||||||||| |||||  |||||||||| ||||||||||||||
Db   99570 AGTCTGAGAGCAATGTGTGTGCTCAAGGTTCAAGGAGGCTTCAAGGAGTAGCAAACCTCCACT 99511

Qy    1479 TTCCAAGTGTATATGTGAGAGAACAGAAAAATATTATAGAAGAAAAAACCCACAGTAGTGAAATC 1538
           ||||||| || ||  || | || |||||| |  ||| ||||||||||| || ||||||||| ||
Db   99510 TTCCAAGCCTATTATTATAGAAGAAGTAAGTAGTGAAAAGAAATTTCCAGTCTGTGAAATC 99451

Qy    1539 TTTACCATCAA 1549
           |||| |||||
Db   99450 TTTATCAGCAA 99440
```

```
TITLE       Wetherby,K.D.; Wiggins,L.; Young,A.; Zhang,L.-H. and Green,E.D.
JOURNAL     NISC Comparative Sequencing Initiative
REFERENCE   Unpublished
AUTHORS     2 (bases 1 to 119266)
            Green,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (01-FEB-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3 (bases 1 to 119266)
AUTHORS     Green,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (24-JUL-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jul 24, 2002 this sequence version replaced gi:18464076.
            -------------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_zoo@nhgri.nih.gov
            ----------- Project Information
            Center project name: cfu
            Center clone name: 299K23

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ----------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 118977 bases at least Q40
            Consensus quality: 119103 bases at least Q30
            Consensus quality: 119162 bases at least Q20
            Insert size: 118000; agarose-fp
            Insert size: 119166; sum-of-contigs
            Quality coverage: 11.69x in Q20 bases; agarose-fp
            Quality coverage: 11.58x in Q20 bases; sum-of-contigs

            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1      78597: contig of 78597 bp in length
            *   78598      78697: gap of unknown length
            *   78698  119266: contig of 40569 bp in length

FEATURES             Location/Qualifiers
     source          1..119266
                     /organism="Felis catus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9685"
                     /clone="RP86-299K23"
                     /clone_lib="RP86"
     misc_feature    1..78597
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    78698..119266
                     /note="assembly_fragment
                     clone_end:T7
```

```
     misc_feature            vector_side:right"
                             87940..119266
                             /note="clone overlaps with GenBank Accession Number
                             AC105439 clone RP86-515N8 (center project name awl)"

ORIGIN

     Query Match             50.9%;   Score 788.4;   DB 2;   Length 119266;
     Best Local Similarity   77.2%;   Pred. No. 2.8e-139;
     Matches 1148;   Conservative   0;   Mismatches 276;   Indels 64;   Gaps 13;
```

```
Qy    77 GGTATCAGAATGGCAAGCAACTTAACCTTCCAGAAAATACAGATGATTGGACAAAAGAG 136
         ||||||||||||||||||||||||||||||||| || |||||||||||||| ||| ||
Db  8288 GCTATCAGAATGGCAATGCCACTGAACCTACCAGAAAATACAGATGACTGAACAGA---  8232

Qy   137 GATGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAGACACAGGGAAATTTGACT 196
         || ||| ||| ||||||||||||||||||||| |||||  |||||||||||||||||
Db  8231 GATATAAATTGATGTTTGGAAAGTCATAGAATTAATCAAAAGCACAGGGAAATTCTTGACT 8172

Qy   197 GAACAAGACGTGAATGGAGCAGTCTT-GAAGTGGTTAAAAAAAGAACATCTTGTTGATAT 255
         |||||||| ||||| || ||||||||  ||||||||  || |||||||| |||||||||
Db  8171 GCAAAAGATATGCATAGATAATCTTGGAAGTGGTTAACGAAAAATCACCTTGTTGATAT 8112

Qy   256 GGGCATCACACATGGACCAGCTATTCAAATAGAAGAACTATTCAAAGAATTGCGGAAAAC 315
         |||||||||| |||||  |||| || || |||||| |||||||||| ||||||||||||
Db  8111 GGGCATAGCACATGGGATGAGCTATCCATATAGAACACAGCTATTCAAATAATTGCAGAAAAC 8052

Qy   316 AGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCCAGTAAAAATGCTCCTAA 375
         ||| ||  |||||| ||||||||||||||||||  || |||||||||||||||||||||
Db  8051 ATCCTCTGAAGATCTTAGTCAGATCGACATGAAAGA-GGGAAAAGGTAGTAAAAATGTTCCTAA 7993

Qy   376 AGACCAAACTGTGTCTCAAAAGAACGTAGAGAAACTTCAAGCAAGAAACTTCAAAGGGTAA 435
         || ||||||||  | || |||||| |||||||||| ||||||||||| |||||||||||
Db  7992 AATACAACATTTGATGTAAGAGAGAAATGGACAAAGCTTCAAGCAAAGCAAAATG--AT 7935

Qy   436 AGAGAACCCAGATATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTC 495
         ||||||||||  ||| ||||  ||| ||||||| |||||||||| ||||||||    |
Db  7934 AGAGAAATCAGTTATGGCTAATGTCTTCTACAGTGGGGTAC-------------AGTC 7892

Qy   496 ACTAAAAGTTGAGCTCATAGAAGATAAAATAGATTATCAAAGGAAAGGCAACCATCCAT 555
         || |||| ||||||||||  || | ||  ||||||||||||||||| ||||||||||
Db  7891 ACTAAAAACTGAGTTCATGGAAGATGAAATAGATGACAAAGGAAAGGCAGCCATCCAG 7832

Qy   556 AGACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTT 615
         |||||||||||||| ||||| ||||||||||||||||||||  || |||  |||||
Db  7831 AGAACTGACATGTGTATCATACTTTTTTGATGAATT-TGTAACCCATATCACTACAAATC 7773

Qy   616 GGATTTAGTCACAGCCTGAAACAGG---ACCAGGCAAATCTCATTGATCCGATACATGA 672
         |||||| ||||||||| ||||||||   ||||| ||||||| ||||| || ||||
Db  7772 AAATTTTAGTTCTACAGCCTGAAATAGGACCACCACTCAATCTCACTGATCCAACAAATTA 7713

Qy   673 ATTCAAAGCC----TTCACAAATACAGCAACAGCAGCCACGAAGAGGATGTCAAGATGAAAT 728
         |||||||||     |||| |||||| |||||  ||||||||| || ||||| ||||||||
Db  7712 ATTCAAAACCTATTGTGAAAATAGAGAAAACAGCTGCAGAAGA-GATGTTAAGGTGAAAT 7654

Qy   729 TTAGCAATGAGGTTTCCGATTTGCTTCAGCTTGTATG--AATTCACGTACCAATGGCAC 786
         |||||||||||| ||| |||||||||| |||||||||  ||||||||| |||||||||||
Db  7653 TTAGCAATGAAGTTTTCCAATTTGCTTCGGGTGTATGATGAAAATTCACCTACTAATGGCAC 7594

Qy   787 TATTCATTTGGAGTCAAAGACAAACCCATGGGAAAATTGTTGGCATCAAAGTCACCAA 846
         ||||||||||||||||||| |||| ||||||||||||||| ||||||||||||||||||
Db  7593 CATTCATTTGGAGTTAAAGGCAAACCCTATGGAAAAAATTCTTTACATGAAAGTCACCAA 7534

Qy   847 TGATACCAAGGAGAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTGAAGA 906
         |   ||||| || || || |||||||||||| || |||||||||||||||||||||||||
Db  7533 TTTCACCAAGGAAGCCCTCATTGACCATTTTAAATTGATGATCATCCACCAGTATTTGAAGA 7474

Qy   907 CCATCAAGTCCACAAGCAAAGAAGTGCATTCGAGAGGCCAAGATTTGTGGAAGTTTTACT 966
         |||||||| ||||||||||| |||  ||||| |||||||||||||| ||| || ||||||
Db  7473 CCATCAGGTCCAAAAAGCCAAAAAGCAAAGAAGTACATTTGAGAGCCAAGATTGTGAGACGTTTTACT 7414

Qy   967 GCCAAATAGTACTCATCTGACAGATTTG-TATTGAAGCAAGGTCATTACAAAGTACTC 1026
         || ||| |||||||| |||||||||||  |||||||||||||| || |||| ||| ||
Db  7413 GCTAAACCTGAGGTATCCTGACAGATTTG-TATTGAAGCAAGGTCATTACAAAGTACTC 7355
```

COMMENT   On Aug 7, 2002 this sequence version replaced gi:18958663.
            ----------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_zoo@nhgri.nih.gov
            ----------------- Project Information
                Center project name: avu
                Center clone name: 410018

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contig have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ----------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 225008 bases at least Q40
            Consensus quality: 225249 bases at least Q30
            Consensus quality: 225423 bases at least Q20
            Insert size: 180000; agarose-fp
            Insert size: 225885; sum-of-contigs
            Quality coverage: 11.24x in Q20 bases; agarose-fp
            Quality coverage: 8.97x in Q20 bases; sum-of-contigs
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *   1   44668: contig of 44668 bp in length
            *   44669   44768: gap of unknown length
            *   44769   72218: contig of 27450 bp in length
            *   72219   72318: gap of unknown length
            *   72319   94001: contig of 21683 bp in length
            *   94002   94102: gap of unknown length
            *   94102 196623: contig of 102522 bp in length
            *   196624 196723: gap of unknown length
            *   196724 225885: contig of 29162 bp in length

FEATURES   Location/Qualifiers
     source   1..225885
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9615"
                     /clone="RP81-410018"
                     /clone_lib="RP81"
     misc_feature   1..118000
                     /note="clone overlaps with GenBank Accession Number
                     AC110665 clone RP81-144D11 (center project name avt)"
     misc_feature   1..44668
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature   44769..72218
                     /note="assembly_fragment"
     misc_feature   72319..94001
                     /note="assembly_fragment"
     misc_feature   94102..196623
                     /note="assembly_fragment"
     misc_feature   144435..225885

```
Qy   1027 TGAATGCCAATATGATACTTCCAGATTAAAATGCAAAATTACAACACAAAATATGGGA 1086
          ||||||| |||||| ||||||||||||||| |||||||||| || ||||||||||||
Db   7354 TGAATATGAACAAGATTATTTCCATACTAAAATGCAAAATCACAACACAGAACATGGAA 7295

Qy   1087 ACAAAGTAAAAAAATTCCACTATTTGTGCGAGATGGGACCAGCTCAAGGACATTACGAA 1146
          || ||||  |||||||||||||||||||||||||||||||||||||||||||||||||
Db   7294 ACTAAGTCCAAAATAATTCAGTCTTCATGTGAGCATGGGCCAGCTCTAAAGACATCATGAA 7235

Qy   1147 AAATAAAGTTGATTTCAGAGCATTTAAAGCAGATTTTAAAACACTGCAGAGTCCAGAAA 1206
          ||||||| ||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   7234 AAATAATGCAGATTTCAGAGCATTTAAATTAGATTTGAAAGCACTGTCAGAGTCTAGAAA 7175

Qy   1207 AGAGCAGAAGAAAAATTCAGAGCAAAATCAAATAAAAAAAAGAAAGAGAGGGACCAAAGTT 1266
          | |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   7174 AGAGGCAGAAGAAACATGCAAAGTAGAAAACAAATAAAAAAAAGAGTAAGGGACCAACACT 7115

Qy   1267 GGTTAAATTATTGACAGGAAATCAAGATTTGTTAGATAATTCATACTATGAAC------A 1320
          ||| ||||||||||||||||| |||||||| ||||||||||||||||||||||      |
Db   7114 TGCTAAATTATTGACAGGAAATCAGGATTTGTTTGATAATTCATTCAAACCACGACTGGT 7055

Qy   1321 GTACATTCTTGTGACAAATAAATGCCACCCGAGTCAACAAAACACTTAGATTTCCTGAA 1380
          |||||||| ||||| |||||||||||||| ||||||||||||||| || || |||||||
Db   7054 ACACATTCTTGTGACAAATAAATGCCATCTGAATCAAAATAAAAACACTTAGATTCCCTAAA 6995

Qy   1381 GGAAATTAAATGGTTTGCTGATTGGAGTTGATCCTGAGTCTAACATCAATGGAGTGGT 1440
          |||||| |||| ||||||||| ||||| |||||| |||||||||||||||||||||||
Db   6994 GGAAATTTAATGACTTGCTGTGTGGGAGTTTAATCCTAAAATCTGTGAGAGAGGGGGTAGT 6935

Qy   1441 CAAAGCTTACAAAGAAGCCGAGTAGCAAACCTTCACTTTCCAAGTGTATATGTAGAACA 1500
                    |||||||||||||||               |||||||||||||||||||
Db   6934 CAAAGATTACAAAGAAG----------------------CCAAATCTGTATGTGTGAGGG 6897

Qy   1501 GAAAACCACCAAATGAGACGATTTCTACTCTAAATCTTTACCATCA 1548
          | ||| |||||||||||| ||||| |||||| |||||||||||| ||
Db   6896 GGAAAAAATACAAAATACAAGAGATTTCTAGTCTGAATCTTTATCAACA 6849
```

RESULT 15
AC113236/c
LOCUS   AC113236   225885 bp   DNA   linear   HTG 07-AUG-2002
DEFINITION Canis familiaris clone RP81-410018, WORKING DRAFT SEQUENCE, 5
            ordered pieces.
ACCESSION   AC113236
VERSION   AC113236.2 GI:22128671
KEYWORDS   HTG; HTGS PHASE2; HTGS_DRAFT.
SOURCE   Canis familiaris (dog)
  ORGANISM Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1 (bases 1 to 225885)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Pagnirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE   NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 225885)
  AUTHORS   Green,E.D.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-FEB-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3 (bases 1 to 225885)
  AUTHORS   Green,E.D.
  TITLE   Direct Submission
  JOURNAL   Submitted (07-AUG-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA

```
                  /note="clone overlaps with GenBank Accession Number
                  AC110666 clone RP81-156E11 (center project name avv)"
misc_feature      196724..225885
                  /note="assembly_fragment
                  clone_end:SP6
                  vector_side:right"
misc_feature      223792..225885
                  /note="clone overlaps with GenBank Accession Number
                  AC110671 clone RP81-69D24 (center project name avv)"
ORIGIN

Query Match       49.0%;   Score 759;   DB 2;   Length 225885;
Best Local Similarity  77.2%;   Pred. No. 8.8e-134;
Matches 1143;   Conservative   0;   Mismatches 285;   Indels 53;   Gaps 16;

Qy      77 GGTATCAGAATGGAAAGCAACTTAACCTTCCAGAAATACAGATGATTGGACAAAAGAG 136
           ||||||||||||||  ||| ||||||||| ||||||||||||| ||  ||||||||||||
Db  125898 GCTATCAGAATGACAGCACCACTGAACCTGCCAAAAATTACAAATGACTGGGCAAAAGAA 125839

Qy     137 GATGTAAATCAGTCGTTAGAAAGTCATAAGAAGTGGTAAAAAACACAGGGAAATTTTGACT 196
           ||||||||||| ||| ||||  |||||||  |||| ||||||||  |  ||||||||
Db  125838 GATGTAAATAGATGGTTGGAAAGTCATAGAATTGATCAAAAATATAGGAAAAACCTTGCCT 125779

Qy     197 GAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAAGGAACATCTTGTTGATATG 256
           ||||||| |||||||| ||  ||||||||||| ||  |||||  |||||| ||||||||
Db  125778 GCAAAAGATGCACATGGAGCAATCTTGAAGTGGT-----AAAATCACCTTGTTGATTTG 125724

Qy     257 GGCATCACACATGACCAGCTATTCAAATAGAGAAACTATTCAAAGAATTGGGAAAACA 316
           ||||||||||||||  |  |||||                 ||||| ||||||||||||
Db  125723 GGCATATCACATGACACATTC-------------AACCTATTCAAGATTTACAGAAAACA 125678

Qy     317 GCCATTGAAGATTCGATTCAGACATCTAAGATGGGGAAGCCCAGTAAAAATGCTCCTAAA 376
           |||||||||||| || || ||||   |||||||||||||| ||||||||||| ||||||
Db  125677 TCTTCTGAGAGATCCTATTCAGACATGTAAGCATAAGGATAAGAGG-GTAGTAAAAATGTTCCTAAA 125619

Qy     377 GACCAAACTGTGTCTCAAAAGGAACGTAGAGAAACTTCAAAGCAAACAAAAGGGTAA 436
           |||||||||||| ||||||||  |||| || |||||||||||||||||| ||||
Db  125618 ATACAACCTTTATGCAAGAGAAAATGGAGAAATGGAGCAAAACAAAATGATAAA 125559

Qy     437 GAGAACCCAGATAATGGCTAATCCGTCTGCAATGAGTACAACTGCTAAGGTTCTAAGTCA 496
           ||||||| || ||||||||||||| || || ||  ||||||||||||| ||||||||||
Db  125558 GAGAAATGAGTCTATGGAAGTGAAGAGTGAAGGTACTAAGGTTCTGCAGCCA 125499

Qy     497 CTAAAAGTTGAGTCATCATAGAAGAGATAAAATAGATTATACAAAGAAAGGCAACCACCATCCATA 556
           ||||||||| ||| ||||||  || |||||||||||| ||||||||| || || ||||||||||
Db  125498 CTAAAAAACTAAGTTCATGGAAGATGAAATAGCTGCACAAAGAAAATCAGCCATCCAGA 125439

Qy     557 GACCTGACATGTGATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAGTTG 616
           |||||||| || |||||||||  ||||  |||| || |||||| ||  |||||||
Db  125438 GAATTAAACATGTGTATCATA-CTTTTGATGAATTC-GTGACTCATATCACTACAAATTG 125381

Qy     617 GATTTTAGTCTACAGCCTGAAACAGG---ACCAGGCAATCTCATTGATCCGATACATGAA 673
           |||  |||||||||  |||||||||||   |||||||| |||||| ||||||||||||||
Db  125380 AGTTTTAGTCTGCATCCTGAAACAGGACCACCACCACTCAACTTCATTGATCCAATACATGAA 125321

Qy     674 TTCAAAGCCTCACAAATACAGCAGCCACAGCAACAGAGAGGATGTCAAGATGAAATTTAGC 733
           |||||||||  ||||||       |||||||||  |||||||| ||||||  ||||||||
Db  125320 TTCAAAGCCTTCACAAAT-------AGCACAGAA-AAAATGTTAAGGTGAAATTTAGC 125270

Qy     734 AATGAGGTTTTCCGATTTGCTTCAGCTTGTATG-AATTCACGTACCAAATGCACTATTCA 792
           ||||  ||| || ||| ||||| ||||     |  |||||||||| ||||||| |||||||
Db  125269 AA-CAAATTTTTCAATTTTTCTTCAGCTTGTATGAAATTCACCTACTAAATGGCACCATTCA 125211

Qy     793 TTTTGGAGTCAAAGCAACAAACCCCATGGGAAAGATTTGTTGGCATCAAATGCATCCAATGATAC 852
           |||||||||| ||||||||||  ||||  |||||||| ||| ||||| ||||| |||||||
Db  125210 TTTTGGAGTCAAAGGCAACAAACCCCATGGGAAATGAAACGTGTTGACATGAAAGTCACCAATTGCAC 125151

Qy     853 CAAGGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAACAGTATTTTGAAGACCATCA 912
           ||||| ||||| | | || |  |||||||| ||||||| || ||||||||||||||||||||
Db  125150 CAAGAAAGCCCTTACTGACCGATTCAAACTGATCCCACTAGTATTTTTGAAGACCATCA 125091

Qy     913 AGTCCAACAAGCAAAGAAGTGCATTCGAGAGCCAAGGATTTGTGGAAGTTTTACTGCCAAA 972
           |||||| ||||||||||  || |||   |||||||| ||||| ||| ||||||||||||||
```

```
Db  125090 GGTCCAAAAAGCAAAGAAGTACAATTGAGAGCCAAGGTTGTTGTAGACATTTTACTGCCGAA 125031
           |||  ||||||||||||   || |||||||| ||||| || ||||||||||||||||||||
Qy     973 TAGT--ACTCTATCTGACAGATTTGTTATTGAAGTGGACATTATTCCACAGTTCTCTGAA 1030
           |||    | ||| ||||||||||||||||||||||| |  ||||||| |||| ||||||||
Db  125030 TAGTAAAATCTGTCTGACAGATTTGTTATTGAAGTGAATGTCATTCCAAAGTACTCTGAA 124971

Qy    1031 TGCCAATATGATTACTTCCAGATTAAAATGCAAAATTACAACACAACAAAATATGGGAACAA 1090
           ||  || ||||| | ||| ||| ||||| | ||||| ||||||||||| |||  |||||||
Db  124970 TGTGAACATGACATTTCTAGACTAAGATACAAAATCACACAACACAGGGAAAACAA 124911

Qy    1091 AGTAAAAAATTCTCACTATTTGTGCGAGATGGGACCAGCTCTAAGGACATTACGAAAAAT 1150
           | | |||| ||||| || ||||||| ||||||| ||| |||||||||| | ||| |||||
Db  124910 AATCCAAAACATTTAGTCTTCATGCCAAGATGGGGCCAACTCTAAGGACACATGATGAAAAAT 124851

Qy    1151 AAAGTTGATTTCAGAGCATTTAAAGCAGATTTAAACACTGGCAGAGTCCAGAAAAGCA 1210
           || |||||||||||||||||||||   ||||||||  || || |||  |||||||| |
Db  124850 AATGAAGATTTCAGAGCATTTAAATTAGATTTAAAGCACTAGCAG-GTGTAGAAAAGAG 124792

Qy    1211 GCAGAAGAAAATTCAGAGCAAATCAGAGTAAAAAACAAAATAAAAGAGAGGACCAAAGTTGGTT 1270
           |||||||||| || |||  ||||   |||||||  |||| | |||| | |||||  | |||
Db  124791 GCAGAAGAAAAGCAAAGTGAAA--AAATAAAAAGGAGAGTAAGGGATGAAGAACACTTGTT 124734

Qy    1271 AAATTATTGACAGGAAAATCAAGAGTTGTTAGATAATTCATACTATGAACAGT--ACATTC 1328
           ||||| |||||||  ||||||  |||||||  ||||| |||| ||  ||| |   ||||||
Db  124733 AAATTGTTGACAGAAAATCAGGATTTGTTGTTGATAATATATCCACAACTGGTACACATTC 124674

Qy    1329 TTGTAACAAATAAATGCCACCCAGATCAAACAAAACACTTAGATTCCTGAAGGAAATTA 1388
           ||||||||||||| | |||| |||||| | |||||   |||| |||||| ||||||||
Db  124673 TTGTAACAAATAAT-CCATC-----CAAATGAAACACCTAGATTTCCTAAAGGAAATTA 124620

Qy    1389 AATGGTTTGCTGTATTGGAGTTTGATCCTGAGTCTAACATCAATGGAGTGGTCAAAGCTT 1448
           ||| |||||||| ||||||||||  |||  ||||| |||  ||  |  |||  ||||||||
Db  124619 AAATGTTTGCTGCTATTGGAGTTTAATCCCAAATCTGTAAGCAAAG----GGGCAAAGCTT 124564

Qy    1449 ACAAAGAAAGCCGAGTAGCAAACCTTCACTTTCCAAGTGTATATGTAGACAGAAACCA 1508
           |||||||||||| |||||||||||||||||||||||||| |||| || |||  |||||
Db  124563 GCAAAGAAAGCCAAGTAGCAAACCTTCACTTTCCAAGTCTATCTGTGGAGGGGAAAACTA 124504

Qy    1509 CACCAAAATGAGACGAGATTTCTTCACTCTAAATCTTTACCATCAA 1549
           ||  ||||||| || |||||||| | ||| ||| ||||| |||||
Db  124503 CATAAAAATGCACAGATTTCTCAGTCTGCAGAACCTTTATCAGTAA 124463
```

Search completed: April 25, 2004, 18:58:31
Job time : 4090.84 secs

```
              GenCore version 5.1.6
     Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   April 25, 2004, 07:00:48 ; Search time 763.869 Seconds
                                       (without alignments)
                                       16458.288 Million cell updates/sec

Title:           US-10-051-835-16
Perfect score:   421
Sequence:        1 atggagttcctcattctgg........ttctatgtcctttcagcgag 421

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        27513289 seqs, 1491090276 residues

Total number of hits satisfying chosen parameters:      55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  EST:*
            1:  em_estba:*
            2:  em_esthum:*
            3:  em_estin:*
            4:  em_estmu:*
            5:  em_estov:*
            6:  em_estpl:*
            7:  em_estro:*
            8:  em_htc:*
            9:  gb_est1:*
            10: gb_est2:*
            11: gb_htc:*
            12: gb_est3:*
            13: gb_est4:*
            14: gb_est5:*
            15: em_estfun:*
            16: em_estom:*
            17: em_gss_hum:*
            18: em_gss_inv:*
            19: em_gss_pln:*
            20: em_gss_vrt:*
            21: em_gss_fun:*
            22: em_gss_mam:*
            23: em_gss_mus:*
            24: em_gss_pro:*
            25: em_gss_rod:*
            26: em_gss_phg:*
            27: em_gss_vrl:*
            28: gb_gss1:*
            29: gb_gss2:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	169	40.1	961	10	BE796744	BE796744 601587368
	2	68.8	16.3	533	28	AQ418633	AQ418633 RPCI-11-2
c	3	58.4	13.9	686	10	BB901475	BE901475 601674934
c	4	56.6	13.4	447	12	BI013264	BI013264 PM4-ET015
c	5	56.6	13.4	598	9	AL549589	AL549589 DKFZp686I
c	6	56.6	13.4	635	13	BX504960	BX504960 DKFZp686I
c	7	56.6	13.4	766	12	BI460285	BI460285 603201916
c	8	56.6	13.4	906	14	CF994492	CF994492 AGENCOURT
c	9	56.6	13.2	990	13	BX334700	BX334700 BX334700
c	10	55.6	13.2	498	28	AQ624094	AQ624094 HS_2100_A
c	11	55.6	13.2	518	28	AQ145903	AQ145903 HS_2219_B
c	12	55.6	13.2	765	28	B2600533	B2600533 WHACG07TF
c	13	55.6	13.2	872	12	BI911783	BI911783 603065205
c	14	55	13.1	640	28	AQ108606	AQ108606 CIT-HSP-2
c	15	55	13.1	645	28	AQ077228	AQ077228 CIT-HSP-2
c	16	55	13.1	676	9	AV730039	AV730039 AV730039
c	17	55	13.1	682	9	AV731500	AV731500 AV731500
c	18	54.6	13.0	495	28	AQ235894	AQ235894 HS_2013_B
c	19	54.6	13.0	554	9	AL705106	AL705106 DKFZp686H
c	20	54.6	13.0	795	12	BG698619	BG698619 602658423
c	21	54.6	13.0	930	13	BQ689079	BQ689079 AGENCOURT
c	22	54.2	12.9	365	28	B30587	B30587 HS-1002-B2-
c	23	54	12.8	214	10	BE142974	BE142974 MRO-HT015
c	24	54	12.8	339	28	AQ566892	AQ566892 HS_2104_B
c	25	54	12.8	341	28	AQ080641	AQ080641 CIT-HSP-2
c	26	54	12.8	348	10	BF920788	BF920788 MR2-NT013
c	27	54	12.8	408	28	B33662	B33662 HS-1023-A1-
c	28	53.4	12.7	401	10	BF759523	BF759523 CM1-CT060
c	29	53.4	12.7	522	12	BG709419	BG709419 60267347I
c	30	53.4	12.7	532	12	BM509480	BM509480 ih19a07-y
c	31	53.2	12.6	620	28	AQ586844	AQ586844 RPCI-11-4
c	32	53	12.6	473	10	AW876755	AW876755 CM2-PT004
c	33	53	12.6	517	28	B49322	B49322 RPCI11-5P11
c	34	53	12.6	590	10	BE144970	BE144970 RC6-HT018
c	35	53	12.6	727	14	CF994332	CF994332 AGENCOURT
c	36	52.8	12.5	651	12	BG718914	BG718914 602699207
c	37	52.8	12.5	701	12	BI559378	BI559378 603253055
c	38	52.8	12.5	833	13	BQ226374	BQ226374 AGENCOURT
c	39	52.6	12.5	320	10	BE143157	BE143157 MRO-HT015
c	40	52.6	12.5	360	12	BX981236	BX981236 RC4-CN008
c	41	52.6	12.5	478	13	BX112097	BX112097 BX112097
c	42	52.6	12.5	483	28	B69221	B69221 CIT-HSP-205
c	43	52.6	12.5	762	28	B90054	B90054 CIT-HSP-216
c	44	52.4	12.4	275	28	B58637	B58637 CIT-HSP-201
c	45	52.4	12.4	403	28	AQ278768	AQ278768 CITBI-E1-

```
RESULT 1
BE796744
LOCUS       BE796744          961 bp    mRNA        linear    EST 20-SEP-2000
DEFINITION  601587368F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3941440 5',
            mRNA sequence.
ACCESSION   BE796744
VERSION     BE796744.1  GI:1021942
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 961)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM793 row: h column: 17
            High quality sequence start: 25
            High quality sequence stop: 689.
```

ALIGNMENTS

FEATURES   Location/Qualifiers
     source   1..961
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3941440"
                     /tissue_type="small cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /cell_line="MGC3"
                     /clone_lib="NIH_MGC_7"
                     /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Query Match 40.1%; Score 169; DB 10; Length 961;
Best Local Similarity 92.2%; Pred. No. 5.4e-37;
Matches 178; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy   18 TGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAACTCTGCGATGCTGTT   77
         |||||||||||||| ||||||||||||||| |||||||||||||||||||||||||||
Db 110 TGGTGGGTTCGTGGTCTCTCACTGACTTCAGGAATGAAGCCACGGACCCCTCGCAATGCTCTT 169

Qy   78 ATGAACTTCCTTAGAATCAACCCATTTGTGAAACACCACATTAATAAGAGCAGTATCTT 137
         |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db 170 ACGAACTTCCTTAGAATCAACCCATTTGTGAAACACCACATTAATAAGAGCAGTATCTT 229

Qy 138 TGGAGATTGGAGAGTCCACCAAGGATGCCCACTGGCTCCCTACAAAGTTTTTATGTGAG 197
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 230 TGGAGATTGGAGAGTCCACCAAGGATGCCCACTGGCTCCCTACAAAGTTTTTATGTGAG 289

Qy 198 GACTCGGTCTCAG 210
         |||||||||||||
Db 290 GACTCGGTCTCAG 302

RESULT 2
AQ418633
LOCUS   AQ418633   533 bp   DNA   linear   GSS 23-MAR-1999
DEFINITION RPCI-11-20304.TV RPCI-11-20304 Homo sapiens genomic clone RPCI-11-20304,
            genomic survey sequence.
ACCESSION   AQ418633
VERSION   AQ418633.1 GI:4476357
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 533)
AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
TITLE   Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
JOURNAL   Unpublished (1997)
COMMENT   Other_GSSs: RPCI-11-20304.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end and search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: T7
Class: BAC ends.

FEATURES   Location/Qualifiers
     source   1..533
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7577907"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-20304"
                     /sex="Male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site 1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

Query Match 16.3%; Score 68.8; DB 28; Length 533;
Best Local Similarity 97.2%; Pred. No. 1.3e-08;
Matches 70; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy   1 ATGGAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAG   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 457 ATGGAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAG   516

Qy   61 AACTCTGCGATG   72
         |||||||||||
Db 517 AACTCTGCGGTG   528

RESULT 3
BE901475/c
LOCUS   BE901475   686 bp   mRNA   linear   EST 29-SEP-2000
DEFINITION 601674934F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3957464 5',
            mRNA sequence.
ACCESSION   BE901475
VERSION   BE901475.1 GI:10390693
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 686)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM835 row: d column: 09
            High quality sequence start: 3
            High quality sequence stop: 686.

FEATURES   Location/Qualifiers
     source   1..686
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3957464"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_21"
                     /note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
                     Site 2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

```
Query Match    13.9%;    Score 58.4;    DB 10;    Length 686;
Best Local Similarity    85.5%;    Pred. No. 1.2e-05;
Matches   65;    Conservative   0;    Mismatches   11;    Indels   0;    Gaps   0;

Qy    3 GGAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAA 62
          ||||||||||| ||||||||| ||||||||||||  |||  ||| ||||| |||||||| |||
Db  258 GGAGTTTCTTCCTTCTGGTGGGGTTCGTGGTCTCGTGTCTCGGCTGGTCTCAGGAGGTGAAGCTGCAGAC 199

Qy   63 CTCTGCGGATGCTGTTA 78
          || |||||| |||| |||
Db  198 CTTTGCGGTGATGTTA 183
```

RESULT 4
BI013264/c
```
LOCUS        BI013264                447 bp    mRNA    linear    EST 13-JUN-2001
DEFINITION   PM4-ET0154-200101-009-c09 ET0154 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BI013264
VERSION      BI013264.1    GI:14417335
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 447)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
             O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
TITLE        Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
PUBMED       10737800
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel.: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=PM4&t2=PM4-ET0154-
             200101-009-c09&t3=2001-01-20&t4=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 446.
FEATURES             Location/Qualifiers
     source          1..447
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="ET0154"
                     /note="Organ: lung_tumor; Vector: puc18; Site 1: SmaI;
                     Site 2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN
```

```
Query Match    13.4%;    Score 56.6;    DB 12;    Length 447;
Best Local Similarity    87.3%;    Pred. No. 3.5e-05;
Matches   62;    Conservative   0;    Mismatches   9;    Indels   0;    Gaps   0;

Qy    2 TGGAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGA 61
          ||||||||| ||||| |||| ||||||||||||  |||  ||| |||||||| |||||||| |||
Db  228 TGGAGTTTCTTCCTTCTGGTGGGGTTCGTGGTCTCGGCTTCCGCTGGTCTCAGGAGTGAAGCTGCAGA 169

Qy   62 ACTCTGCGATG 72
          || ||||||
Db  168 CCTTTGCGAGTG 158
```

RESULT 5
AL549589/c
```
LOCUS        AL549589                598 bp    mRNA    linear    EST 31-MAY-2003
DEFINITION   AL549589 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
             clone CS0DI048YC03 5-PRIME, mRNA sequence.
ACCESSION    AL549589
VERSION      AL549589.2    GI:31271407
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 598)
AUTHORS      Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      On Feb 15, 2001 this sequence version replaced gi:12885723.
             Contact: Genoscope
             Genoscope - Centre National de Sequencage
             BP 191 91006 EVRY cedex - France
             Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
             Library was constructed by Life Technologies, a division of
             Invitrogen. Contact: Feng Liang Email : fliang@lifetech.com URL :
             http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
             Faraday Avenue Genoscope sequence ID : CS0DI048AB02QP1.
FEATURES             Location/Qualifiers
     source          1..598
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI048YC03"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
```

```
Query Match    13.4%;    Score 56.6;    DB 9;    Length 598;
Best Local Similarity    87.3%;    Pred. No. 3.7e-05;
Matches   62;    Conservative   0;    Mismatches   9;    Indels   0;    Gaps   0;

Qy    2 TGGAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGA 61
          ||||||| ||||| ||| |||||||||||||  ||| |||  ||||||||| |||||||| |||
Db  229 TGGAGTTTATTCCTTCGTGGTAGGTTCGTGGTCTCGTGTCTCGGCTGACTTCAGGAGTGAAGCTGCAGA 170

Qy   62 ACTCTGCGATG 72
          || ||||| |||
Db  169 TCTTTGCGGTG 159
```

RESULT 6
BX504960
```
LOCUS        BX504960                635 bp    mRNA    linear    EST 04-SEP-2003
DEFINITION   DKFZp686I13187_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
             DKFZp686I13187 5', mRNA sequence.
ACCESSION    BX504960
VERSION      BX504960.1    GI:32032507
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 635)
```

AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
           Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
           Wiemann,S.
TITLE   EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
           Wellenreuther,R., et al.)
JOURNAL   Unpublished (2003)
COMMENT   Contact: MIPS
           MIPS
           Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
           This is the 5' sequence of the clone insert
           Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
           Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
           sequenced by DKFZ (German Cancer Research Center.
           Heidelberg/Germany) within the cDNA sequencing consortium of the
           German Genome Project.
           No s1 sequence available.
           This clone (DKFZp686I13187) is available at the RZPD in Berlin.
           Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
           Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES   Location/Qualifiers
     source   1..635
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686I13187"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

Query Match   13.4%;   Score 56.6; DB 13; Length 635;
Best Local Similarity   87.3%;   Pred. No. 3.8e-05;
Matches   62; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

Qy   2 TGGAGTTTCTTCATTCGGTAGGTTCGTGGTCCTCTGGCTTCAGGAATGAAGCTGTAGA 61
            |||||||| |||||||||||||||||||| ||||||||||||| ||||||||||| ||
Db   165 TGGAGTTTTATTCCTTCGGTAGGTTCGTGGTTCGTGCTGTCTCGCTGACTTCAGGAGTGAAGCTGCAGA 224

Qy   62 ACTCTGCGATG 72
            || |||||
Db   225 TCTTTGCGGTG 235

RESULT 7
BI460285/c
LOCUS   BI460285   766 bp   mRNA   linear   EST 21-AUG-2001
DEFINITION 603201916F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5267955 5',
            mRNA sequence.
ACCESSION   BI460285
VERSION   BI460285.1 GI:15250941
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 766)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11675 row: p column: 04
            High quality sequence stop: 701.

FEATURES   Location/Qualifiers
     source   1..766
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5267955"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_97"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.2 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

Query Match   13.4%;   Score 56.6; DB 12; Length 766;
Best Local Similarity   87.3%;   Pred. No. 4e-05;
Matches   62; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

Qy   2 TGGAGTTTCTTCATTCGGTAGGTTCGTGGTCCTCTGGCTTCAGGAATGAAGCTGTAGA 61
            |||||||| |||||||||||||||||||| ||||||||||||| ||||||||||| ||
Db   242 TGGAGTTTCTTCCTTCCTTCGGTGGTGGTTCGTGCTGTCTCGCTGGCTTCAGGAGTGAAGCTGCAGA 183

Qy   62 ACTCTGCGATG 72
            || |||||
Db   182 CCTTCGCGGTG 172

RESULT 8
CF994492/c
LOCUS   CF994492   906 bp   mRNA   linear   EST 25-NOV-2003
DEFINITION AGENCOURT 1562-259 NIH_MGC_147 Homo sapiens cDNA clone
            IMAGE:30520429 5', mRNA sequence.
ACCESSION   CF994492
VERSION   CF994492.1 GI:38510552
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 906)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Stefan Hansson
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
            and advice from Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM593 row: 1 column: 14
            High quality sequence stop: 564.

FEATURES   Location/Qualifiers
     source   1..906
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30520429"
                     /tissue_type="Human Placenta"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_147"

```
Db          174 TCTTTGCGGTG 164

RESULT 10
AQ624094/c
LOCUS       AQ624094                498 bp    DNA     linear     GSS 16-JUN-1999
DEFINITION  HS_2100_A1_B10_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2100 Col=19 Row=C, genomic survey
            sequence.
ACCESSION   AQ624094
VERSION     AQ624094.1  GI:5086486
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 498)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
PUBMED      10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2100  row: C  column: 19
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 498.
FEATURES             Location/Qualifiers
     source          1..498
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2100 Col=19 Row=C"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

Query Match             13.2%;  Score 55.6;  DB 28;  Length 498;
Best Local Similarity   87.1%;  Pred. No. 6.9e-05;
Matches    61;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          3 GGAGTTTCTTCATTCTGGTAGGTTCGTGCGTTCTCTGGCTTCAGGAATGAAGCTGTAGAA 62
              ||||||||| ||||||||||||||||||| ||||||||||| ||||||||||||||||||
Db        272 GGAGTTTCTTCTTCTTCCTTCGTGCGTTCGTGCGTTCTCACTGGCTTCAGGAGTGAAGCTGCAGAC 213

Qy         63 CTCTGCGGATG 72
              |||| |||| |
Db        212 CTCCGGTGGTG 203

RESULT 11
AQ145903/c
LOCUS       AQ145903                518 bp    DNA     linear     GSS 26-SEP-1998
DEFINITION  HS_2219_B2_G02_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2219 Col=4 Row=N, genomic survey
            sequence.
ACCESSION   AQ145903
VERSION     AQ145903.1  GI:3536556
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
```

```
                     /note="Organ: placenta; Vector: pBluescriptR; Site_1:
                     alI-XhoI; Site 2: BamH; Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: This is
                     a NIH_MGC library."
ORIGIN

Query Match             13.4%;  Score 56.6;  DB 14;  Length 906;
Best Local Similarity   87.3%;  Pred. No. 4.1e-05;
Matches    62;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 TGGAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTGGCTTCAGGAATGAAGCTGTAGA 61
              |||||||| || |||||||||||||||||||| ||||||||| |||||||||||||||
Db        211 TGGAGTTTATTCCTTCTCGGTAGGTTCGTGGTTCGTGGTCTCGCTGACTTCAGGAGTGAAGCTGCAGA 152

Qy         62 ACTCTGCGGATG 72
              || ||||| ||
Db        151 TCTTTGCGGTG 141

RESULT 9
BX334700/c
LOCUS       BX334700                990 bp    mRNA    linear     EST 01-MAY-2003
DEFINITION  BX334700 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI005YJ22 5-PRIME, mRNA sequence.
ACCESSION   BX334700
VERSION     BX334700.1  GI:30308308
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 990)
AUTHORS     Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9886.f,
            Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fullength.invitrogen.com/ InvitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI005DE11QP1.
FEATURES             Location/Qualifiers
     source          1..990
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI005YJ22"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Query Match             13.4%;  Score 56.6;  DB 13;  Length 990;
Best Local Similarity   87.3%;  Pred. No. 4.2e-05;
Matches    62;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 TGGAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTGGCTTCAGGAATGAAGCTGTAGA 61
              |||||||| || |||||||||||||||||||| ||||||||| |||||||||||||||
Db        234 TGGAGTTTATTCCTTCTCGGTAGGTTCGTGGTCGTGGTCTCGCTGACTTCAGGAGTGAAGCTGCAGA 175

Qy         62 ACTCTGCGGATG 72
              || ||||| ||
```

```
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 518)
AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and
           Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
           scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE    99380589
PUBMED     10449764
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
           High-Throughput Sequencing Center
           University of Washington
           401 Queen Anne Avenue North, Seattle, WA 98109, USA
           Tel: (206) 616-3618
           Fax: (206) 616-3887
           Email: jwallace@u.washington.edu
           Sequence Tagged Connector
           Plate: 2219 row: N column: 4
           Class: BAC ends
           High quality sequence stop: 518.
FEATURES             Location/Qualifiers
     source          1..518
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2219 Col=4 Row=N"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

Query Match             13.2%;  Score 55.6;  DB 28;  Length 518;
  Best Local Similarity 77.9%;  Pred. No. 6.9e-05;
  Matches  67;  Conservative   0;  Mismatches  19;  Indels   0;  Gaps   0;

Qy      2 TGGAGTTCTTCATTCTGGTAGGTTCGTGGTCTTCTGGCTTCAGGAATGAAGCTGTAGA 61
          |||||||| ||||||| |||||  |||||| ||||||||||||||| |||||||||||
Db    333 TGGAGTTCTTCTTCCTTCTGGTGGGTTCGTGGTCTCGCTGCTGGCTTCAGGAGTGAAGCTGCAGA 274

Qy     62 ACTCTGCGATGCTGTTATGAACTTCC 87
          ||||| ||| ||| | |
Db    273 CCTTGACGTTCCAGGTGAGTGTTACC 248


RESULT 12
BZ600533/c
LOCUS       BZ600533                 765 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHACG07TF Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-14B13, genomic survey sequence.
ACCESSION   BZ600533
VERSION     BZ600533.1  GI:31508995
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 765)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     End-sequence profiling: Sequence-based analysis of aberrant genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
  MEDLINE   22709111
   PUBMED   12788976
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..765
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-14B13"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site 1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
ORIGIN

Query Match             13.2%;  Score 55.6;  DB 28;  Length 765;
  Best Local Similarity 87.1%;  Pred. No. 7.6e-05;
  Matches  61;  Conservative   0;  Mismatches  9;  Indels   0;  Gaps   0;

Qy      3 GGAGTTCTTCATTCTGGTAGGTTCGTGGTCTTCTGGCTTCAGGAATGAAGCTGTAGAA 62
          |||||||||| ||||||||| |||||||||||||||||||||||||||||||||||||
Db    152 GGAGTTCTTCTTCCTTCTGCGCGGGGTTCGTGGTCTCGCTGCTGGCTTCAGGAGGTGAAGCTGCAGAC 93

Qy     63 CTCTGCGATG 72
          ||||||||||
Db     92 CTTTGCGGTG 83


RESULT 13
BI911783/c
LOCUS       BI911783                 872 bp    mRNA    linear   EST 16-OCT-2001
DEFINITION  60306520SF1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5214409 5',
            mRNA sequence.
ACCESSION   BI911783
VERSION     BI911783.1  GI:16175655
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 872)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11538 row: i column: 02
            High quality sequence start: 27
            High quality sequence stop: 866.
FEATURES             Location/Qualifiers
     source          1..872
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5214409"
                     /tissue_type="leukocyte"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_118"
                     /note="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV
                     (destroyed); RNA source leukocytes from anonymous pool of
                     non-activated adult donors. Library is oligo-dT primed
```

```
            and directionally cloned (EcoRV site is destroyed upon
            cloning). Average insert size 1.7 kb, insert size range
            1.2-3.3 kb. Library is normalized and enriched for
            full-length clones and was constructed by C. Gruber
            (Invitrogen). Research Genetics tracking code 027. Note:
            this is a NIH_MGC Library."
ORIGIN

Query Match     13.2%;  Score 55.6;  DB 12;  Length 872;
Best Local Similarity   87.1%;  Pred. No. 7.9e-05;
Matches   61;  Conservative   0;  Mismatches   9;  Indels   0;  Gaps   0;

Qy     3 GGAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 62
         |||||||||||| ||||||||||||||||| |||||||||||||||||| || |||
Db   649 GGAGTTTCTTCCTTCGTGGTGGGTTCGGCTGGCTCGCGCTGGCTTCAGGAGTGAAGCTGCAGAC 590

Qy    63 CTCTGCGATG 72
         |||||||||
Db   589 CTTCGGGGTG 580

RESULT 14
AQ108606/c
LOCUS       AQ108606                640 bp    DNA      linear   GSS 29-AUG-1998
DEFINITION  CIT-HSP-2379I19.TF CIT-HSP Homo sapiens genomic clone 2379I19,
            genomic survey sequence.
ACCESSION   AQ108606
VERSION     AQ108606.1  GI:3485296
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 640)
AUTHORS     Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
TITLE       Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
JOURNAL     Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2379I19.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..640
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2379I19"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

Query Match     13.1%;  Score 55;  DB 28;  Length 640;
Best Local Similarity   85.9%;  Pred. No. 0.00011;
Matches   61;  Conservative   0;  Mismatches  10;  Indels   0;  Gaps   0;

Qy     2 TGGAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTCTGGCTTCAGGAATGAAGCTGTAGA 61
         ||||||||||| ||||||||||||||| || ||||||||||||||||||||||| |||
Db   206 TGGAGTTTCTTCCTTCGTGGGTTCGTCGTCGTGGGTTCGCGTCGTGGGGTCTTCAGGAGTGAAGCTGCAGA 147
```

```
Qy    62 ACTCTGCGATG 72
         ||  ||||||
Db   146 CCTTCGTGGTG 136

RESULT 15
AQ077228/c
LOCUS       AQ077228                645 bp    DNA      linear   GSS 20-AUG-1998
DEFINITION  CIT-HSP-2359F6.TR CIT-HSP Homo sapiens genomic clone 2359F6,
            genomic survey sequence.
ACCESSION   AQ077228
VERSION     AQ077228.1  GI:3438412
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 645)
AUTHORS     Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Shizuya,H., Simon,M. and
            Venter,J.C.
TITLE       Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
JOURNAL     Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research.
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..645
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2359F6"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII;.Site_2:
                     HindIII"
ORIGIN

Query Match     13.1%;  Score 55;  DB 28;  Length 645;
Best Local Similarity   85.9%;  Pred. No. 0.00011;
Matches   61;  Conservative   0;  Mismatches  10;  Indels   0;  Gaps   0;

Qy     2 TGGAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTCTGGCTTCAGGATGAAGCTGTAGA 61
         |||||||||||| ||||||||||||||||| || ||||||||||||||||||||||||
Db   275 TGGAGTTTCTTCCTTCGTGGGTTCGTAGTCTTCGCTGTCTTCAGGAGTGAAGCTGCAGA 216

Qy    62 ACTCTGCGATG 72
         ||  ||||||
Db   215 CCTTCGGATG 205

Search completed: April 26, 2004, 01:59:26
Job time : 767.869 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:15:18 ; Search time 20.8853 Seconds
                                   (without alignments)
                                   11186.557 Million cell updates/sec

Title: US-10-051-835-16
Perfect score: 421
Sequence: 1 atggagtttcttcattctgg.........ttctcatgtcctttcagcgag 421

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277445446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents NA:*
              1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
              2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
              3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
              4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
              5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
              6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	55	13.1	426	4	US-09-621-976-10375	Sequence 10375, A
c	2	54.8	13.0	501	4	US-09-621-976-10548	Sequence 10548, A
c	3	52.6	12.5	579	4	US-09-621-976-1382	Sequence 1382, Ap
c	4	51.6	12.3	356	4	US-09-621-976-9793	Sequence 9793, Ap
c	5	51.6	12.3	391	4	US-09-621-976-7956	Sequence 7956, Ap
c	6	51.6	12.3	466	4	US-09-621-976-8468	Sequence 8468, Ap
c	7	51.6	12.3	481	4	US-09-621-976-8002	Sequence 8002, Ap
c	8	51.6	12.3	481	4	US-09-621-976-8003	Sequence 8003, A
c	9	49.8	11.8	418	4	US-09-621-976-15076	Sequence 15076, A
c	10	49.8	11.8	468	4	US-09-621-976-15059	Sequence 15059, A
c	11	49.8	11.8	486	4	US-09-621-976-14952	Sequence 14952, A
c	12	49.8	11.8	517	4	US-09-621-976-10428	Sequence 10428, A
c	13	49.8	11.8	80246	3	US-09-078-294-4	Sequence 4 , Appli
c	14	49.8	11.8	80595	3	US-09-078-294-3	Sequence 3 , Appli
c	15	47	11.2	443	4	US-09-621-976-10565	Sequence 10565, A
c	16	47	11.2	510	4	US-09-621-976-17134	Sequence 17134, A
c	17	46.2	11.0	398	4	US-09-621-976-10095	Sequence 10095, A
c	18	46	10.9	341	4	US-09-621-976-10120	Sequence 10120, A
c	19	45.4	10.8	432	4	US-09-621-976-12460	Sequence 12460, A
c	20	45.2	10.7	451	4	US-09-621-976-10352	Sequence 10352, A
c	21	45.2	10.7	496	4	US-09-621-976-10365	Sequence 10365, A
c	22	45	10.7	421	4	US-09-621-976-10768	Sequence 10768, A
c	23	45	10.7	488	4	US-09-621-976-986	Sequence 986, App
c	24	45	10.7	494	4	US-09-621-976-635	Sequence 635, App
c	25	44.8	10.6	2381	2	US-08-736-770-4	Sequence 4 , Appli
c	26	44.6	10.6	315	4	US-09-621-976-9027	Sequence 9027, Ap
c	27	44.6	10.6	433	4	US-09-621-976-10177	Sequence 10177, A
c	28	44.6	10.6	440	4	US-09-621-976-12535	Sequence 12535, A
c	29	44.6	10.6	458	4	US-09-621-976-12564	Sequence 12564, A
c	30	44.6	10.6	488	4	US-09-621-976-11085	Sequence 11085, A
c	31	44.6	10.6	510	4	US-09-621-976-9389	Sequence 9389, Ap
c	32	44.6	10.6	1091	4	US-09-422-576D-6	Sequence 6, Appli
c	33	44.6	10.5	1091	4	US-09-422-576D-5	Sequence 5, Appli
c	34	44.2	10.5	482	4	US-09-621-976-1994	Sequence 1994, Ap
c	35	43.6	10.4	128779	4	US-09-497-855A-38	Sequence 38, Appl
c	36	43	10.2	430	4	US-09-621-976-10289	Sequence 10289, A
c	37	43	10.2	442	4	US-09-621-976-16336	Sequence 16336, A
c	38	43	10.2	512	4	US-09-621-976-10927	Sequence 10927, A
c	39	43	10.2	513	4	US-09-621-976-16337	Sequence 16337, A
c	40	43	10.2	572	4	US-09-621-976-10590	Sequence 10590, A
c	41	42.8	10.2	418	4	US-09-621-976-18235	Sequence 18235, A
c	42	42.6	10.1	446	4	US-09-621-976-18867	Sequence 18867, A
c	43	42.6	10.1	1165	4	US-09-422-576D-25	Sequence 25, Appl
c	44	42.6	10.1	1831	4	US-09-422-576D-1	Sequence 1, Appli
c	45	42	10.0	375	4	US-09-621-976-10776	Sequence 10776, A

ALIGNMENTS

RESULT 1
US-09-621-976-10375/c
; Sequence 10375, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10375
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-10375

Query Match   13.1%;   Score 55; DB 4; Length 426;
Best Local Similarity   77.1%;   Pred. No. 1e-08;
Matches   64; Conservative   2; Mismatches   17; Indels   0; Gaps   0;

Qy   4 GAGTTTCTTCATTCGGTAGGTTCGTGGTCTCTCGGCTTCCGGAATGAAGCTGTAGAAC 63
            ||||||| |||||:| |||||||||||||||||||||||||| ||||| ||| | |||
Db   305 GAGTTTCTYCCTTCGGTGGTGGTTCGTGGTTCGCGGCTTCAGGAGTGAAGGTGCAGACC 246

Qy   64 TCTGCGGATGCTGTTATGAACTTC 86
            ||||| ||||| ||| ||| |:||
Db   245 TTTGCGGTGAGTGTTACAGCTTY 223

RESULT 2
US-09-621-976-10548/c
; Sequence 10548, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10548
; LENGTH: 501
; TYPE: DNA

; ORGANISM: Homo sapiens
US-09-621-976-10548

Query Match   13.0%;   Score 54.8; DB 4; Length 501;
Best Local Similarity   79.3%;   Pred. No. 1.3e-08;
Matches 65; Conservative   0; Mismatches 17; Indels   0; Gaps   0;

Qy   4 GAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTCTGGCTTCAGGAATGAAGCTGTAGAAC 63
          ||||||||||||||||||| ||||||||||||||||||||| ||| |||
Db 305 GAGTTTCTTCCTTCTGGTGGGTTCGCTCGTGGGTTCAGGAGTGAAGGTGCAGACC 246

Qy   64 TCTGCGATGCTGTTATGAACTT 85
          ||||| ||| || ||||
Db 245 TTTGCCGTGAGTGTTACAGCTT 224

RESULT 3
US-09-621-976-1382/c
; Sequence 1382, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
SEQ ID NO 1382
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 40..246
; NAME/KEY: sig_peptide
; LOCATION: 40..90
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 9.5
; OTHER INFORMATION: seq FVFLLFVISLAAA/AH
US-09-621-976-1382

Query Match   12.5%;   Score 52.8; DB 4; Length 579;
Best Local Similarity   83.3%;   Pred. No. 6.5e-08;
Matches 60; Conservative   0; Mismatches 12; Indels   0; Gaps   0;

Qy   1 ATGGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTCTGGCTTCAGGAATGAAGCTGTAG 60
          |||||||||||||||| ||||||||||||||||||||| |||||||||
Db 303 ATGGAGTTTCTTCCTTCTGGTAGGTTCGTGGTTCGTGTCCGCTGACTTCAGGAGTGATGCTACAG 244

Qy   61 AACTCTGCGATG 72
          ||| |||||
Db 243 ACCTTTGTGGTG 232

RESULT 4
US-09-621-976-9793/c
; Sequence 9793, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
SEQ ID NO 9793
; LENGTH: 356

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-9793

Query Match   12.3%;   Score 51.6; DB 4; Length 356;
Best Local Similarity   76.8%;   Pred. No. 1.3e-07;
Matches 63; Conservative   0; Mismatches 19; Indels   0; Gaps   0;

Qy   3 GGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 62
          |||||||| |||||| |||||||||||||||||||||||||| |||| |||||||||||
Db 267 GGAGTTTCTTCCTTCTGGTGGGTTTGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGAC 208

Qy   63 CTCTGCGATGCTGTTATGAACT 84
          || ||||| ||| ||||
Db 207 CTTCGCGGTGGCTGTTACAACT 186

RESULT 5
US-09-621-976-7956/c
; Sequence 7956, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
SEQ ID NO 7956
; LENGTH: 391
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 306,313,315
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-7956

Query Match   12.3%;   Score 51.6; DB 4; Length 391;
Best Local Similarity   76.8%;   Pred. No. 1.3e-07;
Matches 63; Conservative   0; Mismatches 19; Indels   0; Gaps   0;

Qy   3 GGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 62
          |||||||| |||||| |||||||||||||||||||||||||| |||| |||||||||||
Db 284 GGAGTTTCTTCCTTCTGGTGGGTTTGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGAC 225

Qy   63 CTCTGCGATGCTGTTATGAACT 84
          || ||||| ||| ||||
Db 224 GTTCGAGGTGAGTGTTACAACT 203

RESULT 6
US-09-621-976-8468/c
; Sequence 8468, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
SEQ ID NO 8468
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-8468

Query Match   12.3%;   Score 51.6;   DB 4;   Length 466;
Best Local Similarity   76.8%;   Pred. No. 1.5e-07;
Matches   63;   Conservative   0;   Mismatches   19;   Indels   0;   Gaps   0;

Qy   3 GGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTGGCTTCAGGAATGAAGCTGTAGAA 62
            |||||||||||||| |||||||||||| |||||||||||||||||||||||||| |||
Db   284 GGAGTTCTTCTTCCTTCGTGGTGGGTTCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGAC 225

Qy   63 CTCTGCGATGCTGTTATGAACT 84
            |||| ||| |||| ||||
Db   224 GTTCGAGGTGAGTGTTACAACT 203

RESULT 7
US-09-621-976-8002/c
; Sequence 8002, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8002
;   LENGTH: 481
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-8002

Query Match   12.3%;   Score 51.6;   DB 4;   Length 481;
Best Local Similarity   76.8%;   Pred. No. 1.5e-07;
Matches   63;   Conservative   0;   Mismatches   19;   Indels   0;   Gaps   0;

Qy   3 GGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTGGCTTCAGGAATGAAGCTGTAGAA 62
            |||||||||||||| |||||||||||| |||||||||||||||||||||||||| |||
Db   284 GGAGTTCTTCTTCCTTCGTGGTGGGTTCGTGGTCTTGCTGGCTTCAGGAGTGAAGCTGCAGAC 225

Qy   63 CTCTGCGATGCTGTTATGAACT 84
            |||| ||| |||| ||||
Db   224 GTTCGAGGTGAGTGTTACAACT 203

RESULT 8
US-09-621-976-8003/c
; Sequence 8003, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8003
;   LENGTH: 481
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 76
;   OTHER INFORMATION: n=a, g, c or t
US-09-621-976-8003

Query Match   12.3%;   Score 51.6;   DB 4;   Length 481;
Best Local Similarity   76.8%;   Pred. No. 1.5e-07;

Matches   63;   Conservative   0;   Mismatches   19;   Indels   0;   Gaps   0;

Qy   63 CTCTGCGATGCTGTTATGAACT 84
            |||| ||| |||| ||||
Db   224 GTTCGAGGTGAGTGTTACAACT 203

RESULT 9
US-09-621-976-15076/c
; Sequence 15076, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15076
;   LENGTH: 418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-15076

Query Match   11.8%;   Score 49.8;   DB 4;   Length 418;
Best Local Similarity   82.6%;   Pred. No. 5.6e-07;
Matches   57;   Conservative   0;   Mismatches   12;   Indels   0;   Gaps   0;

Qy   4 GAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTGGCTTCAGGAATGAAGCTGTAGAAC 63
            ||||||||||||| |||||||||||| |||||||||||||||||| ||||
Db   401 GAGTTTCTTCCTTCGATGGGTTCGTGGTCTTGCTGGTTCGACTTCAGGAATGAAGCTTATAGACC 342

Qy   64 TCTGCGATG 72
            || ||||
Db   341 CTTGTGGTG 333

RESULT 10
US-09-621-976-15059/c
; Sequence 15059, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15059
;   LENGTH: 468
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-15059

Query Match   11.8%;   Score 49.8;   DB 4;   Length 468;
Best Local Similarity   82.6%;   Pred. No. 6e-07;
Matches   57;   Conservative   0;   Mismatches   12;   Indels   0;   Gaps   0;

Qy   4 GAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTGGCTTCAGGAATGAAGCTGTAGAAC 63
            ||||||||||||| |||||||||||| |||||||||||||||||| ||||
Db   401 GAGTTTCTTCCTTCGATGGGTTCGTGGTCTTGCTGGTTCGACTTCAGGAATGAAGCTTATAGACC 342

Qy   64 TCTGCGATG 72

```
Db         |||||||
       341 CTTGTGGTG 333
```

RESULT 11
US-09-621-976-14952/c
; Sequence 14952, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; SOFTWARE: Patent.pm
; SEQ ID NO 14952
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 119
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-14952

Query Match   11.8%; Score 49.8; DB 4; Length 486;
Best Local Similarity 82.6%; Pred. No. 6.1e-07;
Matches   57; Conservative   0; Mismatches 12; Indels   0; Gaps   0;

```
Qy     4 GAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAAC 63
         ||||||||| ||||| | ||  ||||||||  | ||| |||||||||||||| | ||
Db   401 GAGTTTCTTCCTTCTCGATGGGGTTCGTGGTCTTGCTGACTTCAGGAATGAAGCTATAGACC 342

Qy    64 TCTGCGATG 72
         |||||||||
Db   341 CTTGTGGTG 333
```

RESULT 12
US-09-621-976-10428/c
; Sequence 10428, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; SOFTWARE: Patent.pm
; NUMBER OF SEQ ID NOS: 19335
; SEQ ID NO 10428
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-10428

Query Match   11.8%; Score 49.8; DB 4; Length 517;
Best Local Similarity 74.1%; Pred. No. 6.3e-07;
Matches   63; Conservative   0; Mismatches 22; Indels   0; Gaps   0;

```
Qy     4 GAGTTTCTTCATTCTGGTAGGTGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAAC 63
         ||||||||| ||||| |||| | |||||||||||||| | ||||||||| |||| || |
Db   384 GAGTTTCTTCCTTCTCGGTGGGTGGGTTCGTGGTCTCGTGGTCTTCAGGAATGAAACTGCGACC 325

Qy    64 TCTGCGATGCTGTTATGAACTTCCC 88
         ||||||| |
Db   324 TTCAGCATTGAGTGTTACAGCTCTC 300
```

RESULT 13
US-09-078-294-4/c
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match   11.8%; Score 49.8; DB 3; Length 80246;
Best Local Similarity 82.6%; Pred. No. 8.5e-06;
Matches   57; Conservative   0; Mismatches 12; Indels   0; Gaps   0;

```
Qy      4 GAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAAC 63
          ||||||||| ||||| |  ||  |||| ||||||| || ||||||| ||||||||||  |
Db  59664 GAGTTTCTTCCTTCTTCGGTGGGTTTGTGGTCTCCACTGACTTCAGGAGTGAAGCTGCAGAGC 59605

Qy     64 TCTGCGATG 72
          ||| | |||
Db  59604 TTCGCAGTG 59596
```

RESULT 14
US-09-078-294-3/c
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match   11.8%; Score 49.8; DB 3; Length 80595;
Best Local Similarity 82.6%; Pred. No. 8.5e-06;
Matches   57; Conservative   0; Mismatches 12; Indels   0; Gaps   0;

```
Qy      4 GAGTTTCTTCATTCTGGTAGGTTCGTGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAAC 63
          ||||||||| ||||| |  ||  ||| | |||||| || ||||||| |||||||||||  |
Db  59897 GAGTTTCTTCCTTCTTCGGTGGGTTTGTGGTCTCCACTGACTTCAGGAGTGAAGCTGCAGAGC 59838

Qy     64 TCTGCGATG 72
          ||| | |||
Db  59837 TTCGCAGTG 59829
```

RESULT 15
US-09-621-976-10565/c
; Sequence 10565, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.

```
;   APPLICANT:  Jobert, S.
;   APPLICANT:  Giordano, J.Y.
;   TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;   FILE REFERENCE: GENSET.054PR2
;   CURRENT APPLICATION NUMBER: US/09/621,976
;   CURRENT FILING DATE: 2000-07-21
;   NUMBER OF SEQ ID NOS: 19335
;   SOFTWARE: Patent.pm
;   SEQ ID NO 10565
;     LENGTH: 443
;     TYPE: DNA
;     ORGANISM: Homo sapiens
;     FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 356
;     OTHER INFORMATION: n=a, g, c or t
US-09-621-976-10565

Query Match      11.2%;  Score 47;  DB 4;  Length 443;
Best Local Similarity   78.9%;  Pred. No. 5e-06;
Matches   56;  Conservative   0;  Mismatches   15;  Indels   0;  Gaps   0;

Qy      2 TGGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGA 61
          ||||||||||| ||| || ||||||||| |  ||||||||||  || |||||| ||||| |||||
Db    285 TGGAGTTTCTTCCTTCCAGTGGGTTCGTTGGTCTTGGTGACTTCAGGAGTGAAGCCGTAGA 226

Qy     62 ACTCTGCGATG 72
          || || |||||
Db    225 CCTTCGCAGTG 215

Search completed: April 26, 2004, 02:11:09
Job time : 21.8853 secs
```

```
         GenCore version 5.1.6

     Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      April 24, 2004, 23:47:42 ; Search time 111.862 Seconds
                                          (without alignments)
                                          15988.422 Million cell updates/sec

Title:             US-10-051-835-16
Perfect score:     421
Sequence:          1 atggagttcctcattctgg.........ttctatgtccctttcagcgcgag 421

Scoring table:     IDENTITY_NUC
                   Gapop 10.0 , Gapext 1.0

Searched:          3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:      6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    N_Geneseq_29Jan04:*
              1: geneseqn1980s:*
              2: geneseqn1990s:*
              3: geneseqn2000s:*
              4: geneseqn2001as:*
              5: geneseqn2001bs:*
              6: geneseqn2002s:*
              7: geneseqn2003as:*
              8: geneseqn2003bs:*
              9: geneseqn2003cs:*
             10: geneseqn2004s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	421	100.0	421	8	ACD42206	Acd42206 Human HLA
2	160	38.0	481	8	ACH32547	Ach32547 Human end
3	141	33.5	415	4	AAI93591	Aai93591 Human pol
4	56.6	13.4	587	5	AAS90405	Aas90405 DNA encod
5	56.6	13.4	2566	4	AAH18588	Aah18588 Human cDN
6	56.4	13.4	339	5	AAS66944	Aas66944 DNA encod
7	56.4	13.4	648	5	AAS78088	Aas78088 DNA encod
8	56.4	13.4	648	5	AAS89070	Aas89070 DNA encod
9	56.4	13.4	648	5	AAS64261	Aas64261 DNA encod
10	55	13.1	380	3	AAC05498	Aac05498 Human sec
11	55	13.1	1173	5	AAS89076	Aas89076 DNA encod
12	54.8	13.0	465	4	AAI19816	Aai19816 Probe #97
13	54.8	13.0	465	4	AAI10563	Aai10563 Probe #49
14	54.8	13.0	465	4	ABA52210	Aba52210 Human foe
15	54.8	13.0	465	4	ABA64843	Aba64843 Human foe
16	54.8	13.0	465	4	AAI31816	Aai31816 Probe #50
17	54.8	13.0	465	4	AAI45011	Aai45011 Probe #13
18	54.8	13.0	465	4	ABA46962	Aba46962 Human bre
19	54.8	13.0	465	4	ABA22015	Aba22015 Probe #48
20	54.8	13.0	465	4	ABA31959	Aba31959 Probe #10
c 21	54.8	13.0	465	4	AAK25933	Aak25933 Human bon
c 22	54.8	13.0	465	4	AAK39006	Aak39006 Human bon
c 23	54.8	13.0	465	4	AAK13272	Aak13272 Human bra
c 24	54.8	13.0	465	4	AAK00486	Aak00486 Human bra
c 25	54.8	13.0	465	4	ABS38593	Abs38593 Human liv
c 26	54.8	13.0	465	4	ABS25522	Abs25522 Human liv
c 27	54.8	13.0	465	5	AAI00495	Aai00495 Probe #48
c 28	54.8	13.0	465	6	AAI05532	Aai05532 Probe #55
c 29	54.8	13.0	465	6	ABS13090	Abs13090 Human gen
c 30	54.8	13.0	465	6	ABS00516	Abs00516 Human gen
c 31	54.6	13.0	2943	5	AAS80270	Aas80270 DNA encod
c 32	54.2	12.9	558	5	AAS64258	Aas64258 DNA encod
c 33	54.2	12.9	1227	5	AAS72945	Aas72945 DNA encod
c 34	54	12.8	552	5	AAS91453	Aas91453 DNA encod
c 35	54	12.8	1089	5	AAS75340	Aas75340 DNA encod
c 36	54	12.8	1089	5	AAS75103	Aas75103 DNA encod
c 37	54	12.8	1328	5	AAS64807	Aas64807 DNA encod
c 38	54	12.8	237961	6	ABQ80552	Abq80552 Human Can
c 39	53.4	12.7	2612	4	AAH17215	Aah17215 Human CDN
c 40	53	12.6	68355	7	ACF62737	Acf62737 Cancer ba
c 41	53	12.6	68355	7	ADB20852	Adb20852 MRP1 base
c 42	53	12.6	68355	9	ADB87941	Adb87941 Human UGT
c 43	53	12.6	68355	9	ADB96924	Adb96924 Human MDR
c 44	53	12.6	68355	9	ADB92115	Adb92115 Human MDR
c 45	53	12.6	186591	7	ACF62750	Acf62750 Cancer ba

                                                ALIGNMENTS

RESULT 1
ACD42206
ID   ACD42206 standard; cDNA; 421 BP.
XX
AC   ACD42206;
XX
DT   05-SEP-2003   (first entry)
XX
DE   Human HLA class I region cDNA Incyte 428822.1.
XX
KW   Human; ss; DNA methylation; cancer; colon cancer.
XX
OS   Homo sapiens.
XX
PN   US2003013099-A1.
XX
PD   16-JAN-2003.
XX
PF   07-MAR-2002; 2002US-00093766.
XX
PR   19-MAR-2001; 2001US-0277380P.
XX
PA   (LASE/) LASEK A K W.
PA   (JONE/) JONES D A.
PA   (KARP/) KARPF A R.
XX
PI   Lasek AKW, Jones DA, Karpf AR;
XX
DR   WPI; 2003-503249/47.
XX
PT   New combination comprising cDNAs that are expressed in a disorder or
PT   process associated with DNA methylation, useful for diagnosing, staging,
PT   treating or monitoring treatment of cancer, e.g. colon cancer.
XX
PS   Claim 2; Page 23; 66pp; English.
XX
CC   The invention relates to a combination comprising cDNAs which are
CC   expressed in a disorder or process associated with DNA methylation. The
CC   combination and cDNAs are useful for diagnosing, staging, treating or
CC   monitoring treatment of cancer, e.g. colon cancer and for detecting
CC   changes in expression of genes encoding proteins that are associated with
CC   DNA methylation. The protein is useful for screening molecules or
CC   compounds to identify at least one ligand that binds to the protein and
CC   for producing an antibody. The present sequence represents a cDNA
CC   expressed in a disorder or process associated with DNA methylation
XX

SQ   Sequence 421 BP; 99 A; 87 C; 93 G; 142 T; 0 U; 0 Other;

Query Match
Best Local Similarity   100.0%;   Score 421; DB 8; Length 421;
                        100.0%;   Pred. No. 2.1e-122;
Matches   421;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 ATGGAGTTCTCATTCGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAG 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGAGTTCTCATTCGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAG 60

Qy   61 AACCTCGGCATGCGTGTATGAACTTCCCTAGAATCAACCCATTGTGAAACACCACATT 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AACCTCGGCATGCGTGTATGAACTTCCCTAGAATCAACCCATTGTGAAACACCACATT 120

Qy 121 AATAAGACAGTATCTTTGGAGATTGGAGAGTCCACCAAGAGATGCCCACTGGCTCCCTAC 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AATAAGACAGTATCTTTGGAGATTGGAGAGTCCACCAAGAGATGCCCACTGGCTCCCTAC 180

Qy 181 AAAGTTTTTATGTGAGGACTCGGTCTCAGATATGGCTACTGGGATTCTGTTCAAACTGA 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 AAAGTTTTTATGTGAGGACTCGGTCTCAGATATGGCTACTGGGATTCTGTTCAAACTGA 240

Qy 241 GTGAAGACATCATCATCTTTCTTTGAGATCTTCCTTGCTTCTCTGGCCAAAGAAGTTATTC 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GTGAAGACATCATCATCTTTCTTTGAGATCTTCCTTGCTTCTCTGGCCAAAGAAGTTATTC 300

Qy 301 CCTGACCCGACTTTCCAAAGAGTACTAGTTCCTTTTAGCAGTGATTAGAAACCTGATCTG 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 CCTGACCCGACTTTCCAAAGAGTACTAGTTCCTTTTAGCAGTGATTAGAAACCTGATCTG 360

Qy 361 AGTACTACTGTGTGCTTGTGACTAGTGCTTGTGGTTTATTGTTTCTATGTCCTTTCAGCGA 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AGTACTACTGTGTGCTTGTGACTAGTGCTTGTGGTTTATTGTTTCTATGTCCTTTCAGCGA 420

Qy 421 G 421
        |
Db 421 G 421

RESULT 2
ACH32547
ID   ACH32547 standard; cDNA; 481 BP.
XX
AC   ACH32547;
XX
DT   13-OCT-2003 (first entry)
XX
DE   Human endothelial cell cDNA #680.
XX
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2003073623-A1.
XX
PD   17-APR-2003.
XX
PF   30-JUL-2001; 2001US-00918995.
XX
PR   30-JUL-2001; 2001US-00918995.
XX
PA   (DRMA/) DRMANAC R T.
PA   (LABA/) LABAT I.
PA   (STAC/) STACHE-CRAIN B.
PA   (DICK/) DICKSON M C.
PA   (JONE/) JONES L W.
XX
PI   Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful

PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 19759; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present invention
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 481 BP; 138 A; 93 C; 121 G; 126 T; 0 U; 3 Other;

Query Match   38.0%;   Pred. No. 5e-40;
Best Local Similarity   89.6%;   Score 160; DB 8; Length 481;
Matches   172;   Conservative   0;   Mismatches   20;   Indels   0;   Gaps   0;

Qy   18 TGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAACTCTGCGATGCTGTT 77
        ||||| |||| |||||||| | |||||| |||||||||| | | | ||| ||||||||
Db   59 TGGTGGGTTCCTGGTCTCACTGACTTCAAGAATGAAGCCACGGACCCTCGCAATGCTGAT 118

Qy   78 ATGAACTTCCCTAGAATCAACCCATTGTGAAACAACACCATTAATAAGACAGTATCTT 137
        | ||||||||||||||||||||||||||||||||||||||||| | || |||| |||
Db 119 ACGAACTTCCCTAGAATCAACCCATTGTGAAACAACACCATTACTAAGAGCAGAATCTT 178

Qy 138 TGGAGATTGGAGAGTCCACCAAGGAGGATGCCCCACTGGCTCCCTACAAAGTTTTTATGTGAG 197
        |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db 179 TGGAGATTGGAGAGTCCACCAAGGAGGATGCCCCACTGGCTCCCTACAAAGCTTTTTATGTGAG 238

Qy 198 GACTCGGTCTCA 209
        |||||| ||||
Db 239 GACTCGCCTCTCA 250

RESULT 3
AAI93591
ID   AAI93591 standard; cDNA; 415 BP.
XX
AC   AAI93591;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 13651.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US004927.
XX
PR   28-FEB-2000; 2000US-00515126.
PR   18-MAY-2000; 2000US-00577409.
XX

PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AA013660.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT   and treating e.g. leukemia, inflammation and immune disorders.
XX
PS   Claim 1; SEQ ID NO 13651; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 415 BP; 109 A; 99 C; 96 G; 110 T; 0 U; 1 Other;

Query Match   33.5%; Score 141; DB 4; Length 415;
Best Local Similarity 100.0%; Pred. No. 4.6e-34;
Matches 141; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   70 ATGCTGTTATGAACTTCCCTTAGAATCAACCCATTTGTGAAACACCACATTAATAAGAGC 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   221 ATGCTGTTATGAACTTCCCTTAGAATCAACCCATTTGTGAAACACCACATTAATAAGAGC 280

Qy   130 AGTATCTTTGGAGAATGGAGAGTCCACCAAGGATGCCCACTGGCTCCCTACAAAGTTTTT 189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   281 AGTATCTTTGGAGAATGGAGAGTCCACCAAGGATGCCCACTGGCTCCCTACAAAGTTTTT 340

Qy   190 TATGTGAGGACTCGGTCTCAG 210
          |||||||||||||||||||||
Db   341 TATGTGAGGACTCGGTCTCAG 361

RESULT 4
AAS90405
ID   AAS90405 standard; cDNA; 587 BP.
XX
AC   AAS90405;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #26209.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX

DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG26218.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 26209; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 587 BP; 152 A; 122 C; 169 G; 144 T; 0 U; 0 Other;

Query Match   13.4%; Score 56.6; DB 5; Length 587;
Best Local Similarity 87.3%; Pred. No. 2.3e-07;
Matches 62; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

Qy   2 TGGAGTTCTCATTCTGGTAGGTTCGTGGTCTTCTCGGCTTCAGGAATGAAGCTGTAGA 61
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   476 TGGAGTTCTTCCTTCCTTCCTTGTGGGGTTCGTGGTCTCTCTGGCTTCAGGAGTGAAGCTGCAGA 535

Qy   62 ACTCTGCGATG 72
          |||||||||||
Db   536 CCTTCGTGGTG 546

RESULT 5
AAH18588
ID   AAH18588 standard; cDNA; 2566 BP.
XX
AC   AAH18588;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:18776.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0016126.
XX
PR   29-JUL-1999; 99JP-00248036.
PR   27-AUG-1999; 99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX

PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 18776; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides, or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 2566 BP; 512 A; 568 C; 843 G; 643 T; 0 U; 0 Other;

Query Match   13.4%; Score 56.6; DB 4; Length 2566;
Best Local Similarity   87.3%; Pred. No. 4.3e-07;
Matches   62; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

Qy   2 TGGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGA 61
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1124 TGGAGTTCTTCTTCCTTCTGGTGTGGGTTCGTGGTTCGTGGCTTCAGGAGTGAAGCTGCAGA 1183

Qy   62 ACTCTGCGATG 72
              || ||| ||||
Db   1184 CCTTTGCAGTG 1194

RESULT 6
AAS66944
ID   AAS66944 standard; cDNA; 339 BP.
XX
AC   AAS66944;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #2748.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX

PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSB-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG02757.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 2748; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have application in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 339 BP; 45 A; 94 C; 120 G; 80 T; 0 U; 0 Other;

Query Match   13.4%; Score 56.4; DB 5; Length 339;
Best Local Similarity   85.1%; Pred. No. 2.1e-07;
Matches   63; Conservative   0; Mismatches   11; Indels   0; Gaps   0;

Qy   2 TGGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGA 61
              |||||||||||| ||||||||||| ||||||||||||||||||||||||||||||||||
Db   86 TGGAGTTCTTCTTCCTTCTGGTGTGGGTTCGTGGTTCGTGGCTTCAGGAGTGAAGCTGCAGA 145

Qy   62 ACTCTGCGATGCTG 75
              || |||||||||||
Db   146 CCTTTGCGACTCAG 159

RESULT 7
AAS78088
ID   AAS78088 standard; cDNA; 648 BP.
XX
AC   AAS78088;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #13892.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX

```
PF    30-MAR-2001; 2001WO-US008631.
XX
PR   (31-MAR-2000; 2000US-00540217.
PR    23-AUG-2000; 2000US-00649167.
XX
PA   (HYSB-) HYSEQ INC.
XX
PI    Drmanac RT, Liu C, Tang YT;
XX
DR    WPI; 2001-639362/73.
DR    P-PSDB; ABG13901.
XX
PT    New isolated polynucleotide and encoded polypeptides, useful in
PT    diagnostics, forensics, gene mapping, identification of mutations
PT    responsible for genetic disorders or other traits and to assess
PT    biodiversity.
XX
PS    Claim 1; SEQ ID NO 13892; 103pp; English.
XX
CC    The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC    sequences. (I) is useful as hybridisation probes, polymerase chain
CC    reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC    and in recombinant production of (II). The polynucleotides are also used
CC    in diagnostics as expressed sequence tags for identifying expressed
CC    genes. (I) is useful in gene therapy techniques to restore normal
CC    activity of (II) or to treat disease states involving (II). (II) is
CC    useful for generating antibodies against it, detecting or quantitating a
CC    polypeptide in tissue, as molecular weight markers and as a food
CC    supplement. (II) and its binding partners are useful in medical imaging
CC    of sites expressing (II). (I) and (II) are useful for treating disorders
CC    involving aberrant protein expression or biological activity. The
CC    polypeptide and polynucleotide sequences have application in
CC    diagnostics, forensics, gene mapping, identification of mutations
CC    responsible for genetic disorders or other traits to assess biodiversity
CC    and to produce other types of data and products dependent on DNA and
CC    amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC    coding sequences of the invention. Note: The sequence data for this
CC    patent did not appear in the printed specification, but was obtained in
CC    electronic format directly from WIPO at
CC    ftp.wipo.int/pub/published_pct_sequences
XX
SQ    Sequence 648 BP; 126 A; 171 C; 212 G; 139 T; 0 U; 0 Other;

Query Match             13.4%;  Score  56.4;  DB 5;  Length 648;
Best Local Similarity   80.5%;  Pred. No. 2.7e-07;
Matches   66;  Conservative   0;  Mismatches  16;  Indels   0;  Gaps   0;

Qy          3 GGAGGTTCTTCATTCTGGTAGGTTCGTGTCTCTCGGCTTCAGGAATGAAGCTGTAGAA 62
              |||||||||||    |||||||| ||||||| ||||| ||||||| |||||||
Db         87 GGAGTTCTTCCTTCCTTCGTGGGGTTCGTGGTCTCGCTGGGTTCAGGAGTGAAGCTTCAGAC 146

Qy         63 CTCTGCGATGCTGTTATGAACT 84
              ||||||| |||    |||||||
Db        147 CTTTGCGGTGAGTTTTACAGCT 168

RESULT 9
AAS64261
ID    AAS64261 standard; cDNA; 648 BP.
XX
AC    AAS64261;
XX
DT    13-FEB-2002 (first entry)
XX
DE    DNA encoding novel human diagnostic protein #65.
XX
KW    Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW    food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS    Homo sapiens.
XX
```

```
PD    11-OCT-2001.
XX
PF    30-MAR-2001; 2001WO-US008631.
XX
PR    31-MAR-2000; 2000US-00540217.
PR    23-AUG-2000; 2000US-00649167.
XX
PA   (HYSB-) HYSEQ INC.
XX
PI    Drmanac RT, Liu C, Tang YT;
XX
DR    WPI; 2001-639362/73.
DR    P-PSDB; ABG24883.
XX
PT    New isolated polynucleotide and encoded polypeptides, useful in
PT    diagnostics, forensics, gene mapping, identification of mutations
PT    responsible for genetic disorders or other traits and to assess
PT    biodiversity.
XX
PS    Claim 1; SEQ ID NO 24874; 103pp; English.
XX
CC    The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC    sequences. (I) is useful as hybridisation probes, polymerase chain
CC    reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC    and in recombinant production of (II). The polynucleotides are also used
CC    in diagnostics as expressed sequence tags for identifying expressed
CC    genes. (I) is useful in gene therapy techniques to restore normal
CC    activity of (II) or to treat disease states involving (II). (II) is
CC    useful for generating antibodies against it, detecting or quantitating a
CC    polypeptide in tissue, as molecular weight markers and as a food
CC    supplement. (II) and its binding partners are useful in medical imaging
CC    of sites expressing (II). (I) and (II) are useful for treating disorders
CC    involving aberrant protein expression or biological activity. The
CC    polypeptide and polynucleotide sequences have application in
CC    diagnostics, forensics, gene mapping, identification of mutations
CC    responsible for genetic disorders or other traits to assess biodiversity
CC    and to produce other types of data and products dependent on DNA and
CC    amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC    coding sequences of the invention. Note: The sequence data for this
CC    patent did not appear in the printed specification, but was obtained in
CC    electronic format directly from WIPO at
CC    ftp.wipo.int/pub/published_pct_sequences
XX
SQ    Sequence 648 BP; 126 A; 171 C; 212 G; 139 T; 0 U; 0 Other;

Query Match             13.4%;  Score  56.4;  DB 5;  Length 648;
Best Local Similarity   80.5%;  Pred. No. 2.7e-07;
Matches   66;  Conservative   0;  Mismatches  16;  Indels   0;  Gaps   0;

Qy          3 GGAGGTTCTTCATTCTGGTAGGTTCGTGTCTCTCGGCTTCAGGAATGAAGCTGTAGAA 62
              |||||||||||    |||||||| ||||||| ||||| ||||||| |||||||
Db         87 GGAGTTCTTCCTTCCTTCGTGGGGTTCGTGGTCTCGCTGGGTTCAGGAGTGAAGCTTCAGAC 146

Qy         63 CTCTGCGATGCTGTTATGAACT 84
              ||||||| |||    |||||||
Db        147 CTTTGCGGTGAGTTTTACAGCT 168

RESULT 8
AAS89070
ID    AAS89070 standard; cDNA; 648 BP.
XX
AC    AAS89070;
XX
DT    13-FEB-2002 (first entry)
XX
DE    DNA encoding novel human diagnostic protein #24874.
XX
KW    Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW    food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS    Homo sapiens.
XX
```

PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG00074.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 65; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have application in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 648 BP; 126 A; 171 C; 212 G; 139 T; 0 U; 0 Other;

Query Match   13.4%; Score 56.4; DB 5; Length 648;
Best Local Similarity   80.5%; Pred. No.2.7e-07;
Matches   66; Conservative   0; Mismatches   16; Indels   0; Gaps   0;

Qy   3 GGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAA 62
              ||||||| |||||| |||||||||||||||| |||||||||||||||||| |||||||
Db   87 GGAGTTTCTTCCTTCTGGTGGGTTCGTGTGGTCTCGCTGGCTTCAGGAGTGAAGCTTCAGAC 146

Qy   63 CTCTGCGATGCTGTTATGAACT 84
              || ||||| || || |||
Db   147 CTTTGCGGTGAGTTTTACAGCT 168

RESULT 10
AAC05498/c
ID   AAC05498 standard; cDNA; 380 BP.
XX
AC   AAC05498;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 9573.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX

OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999; 99US-0122487P.
XX
PA   (GBST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 9573; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 380 BP; 107 A; 107 C; 80 G; 84 T; 0 U; 2 Other;

Query Match   13.1%; Score 55; DB 3; Length 380;
Best Local Similarity   85.9%; Pred. No. 6.1e-07;
Matches   61; Conservative   0; Mismatches   10; Indels   0; Gaps   0;

Qy   2 TGGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGA 61
              |||||||||| ||||| ||||||||||| || |||||||||||| |||||||||||||
Db   182 TGGAGTTTCTTCCTTCTCGGTGGGTTCGTGTGGGTTCGTCGTGGCTTCGGCTGGCTTCGCTGGCTTCAGGAGTGAAGCTGCAGA 123

Qy   62 ACTCTGCGATG 72
              || ||||
Db   122 CCTTCCCGTG 112

RESULT 11
AAS89076
ID   AAS89076 standard; cDNA; 1173 BP.
XX
AC   AAS89076;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #24880.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX

PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG24889.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 24880; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1173 BP; 204 A; 342 C; 376 G; 251 T; 0 U; 0 Other;

Query Match   13.1%; Score 55; DB 5; Length 1173;
Best Local Similarity 85.9%; Pred. No. 9.8e-07;
Matches   61; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy   2 TGGAGTTCTTCATTCTGGTAGGTTCGTTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGA 61
         ||||||||||| |||| ||| ||||||| | || || ||||| ||||| ||||| ||
Db 170 TGGAGTTCTTCTTCCTTCTGGGTGGGTTCGTTCGTGCTGTCTCGCGCTTCTGGCTTCAGGAGTGAAGCCGCAGA 229

Qy   62 ACTCTGCGATG 72
         || |||||||
Db 230 CCTTCGCGGTG 240

RESULT 12
AAI19816/c
ID AAI19816 standard; DNA; 465 BP.
XX
AC AAI19816;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #9749 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX

PF 30-JAN-2001; 2001WO-US000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-02346872.
PR 27-SEP-2000; 2000US-0363359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX
PS Claim 25; SEQ ID NO 9749; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 465 BP; 121 A; 129 C; 114 G; 101 T; 0 U; 0 Other;

Query Match   13.0%; Score 54.8; DB 4; Length 465;
Best Local Similarity 79.3%; Pred. No. 7.6e-07;
Matches   65; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy   3 GGAGTTCTTCATTCTGGTAGGTTCGTTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAA 62
         ||||||||| |||| ||| ||||||| | ||| || ||||| ||||| ||||| |||
Db 403 GGAGTTCTTCTTCCTTCTGTGGGGTTCGTTCGTGGTCTTGGATGGCTTCAGGAGTGAAGCCTGCAGAC 344

Qy   63 CTCTCGCGATGCTGTTATGAACT 84
         || |||| || |||
Db 343 CTTTGCGGTGAAATGTTACACT 322

RESULT 13
AAI10563/c
ID AAI10563 standard; DNA; 465 BP.
XX
AC AAI10563;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #496 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX
PS Claim 25; SEQ ID NO 496; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 465 BP; 121 A; 129 C; 114 G; 101 T; 0 U; 0 Other;

Query Match   13.0%; Score 54.8; DB 4; Length 465;
Best Local Similarity   79.3%; Pred. No. 7.6e-07;
Matches   65; Conservative   0; Mismatches   17; Indels   0; Gaps   0;

Qy   3 GGAGTTTCTTCATTCTGGTAGGTTCGTGGTCCTCTCGGCTTCAGGAATGAAGCTGTAGAA 62
              ||||||||||| ||||||||| ||||||||| |||||||||||| ||||||| ||
Db   403 GGAGTTTCTTCCTTCTGGTGGGTTCGTGGTCTGGTCGGCTTCAGGAGTGAAGCTGCAGAC 344

Qy   63 CTCTGCGATGCTGTTATGAACT 84
              ||| ||| || |||| ||
Db   343 CTTTGCGGTGAATGTTACAACT 322

RESULT 14
ABA52210/c
ID ABA52210 standard; DNA; 465 BP.
XX
AC ABA52210;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #515.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.

XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.
XX
PS Claim 1; SEQ ID NO 515; 639pp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 465 BP; 121 A; 129 C; 114 G; 101 T; 0 U; 0 Other;

Query Match   13.0%; Score 54.8; DB 4; Length 465;
Best Local Similarity   79.3%; Pred. No. 7.6e-07;
Matches   65; Conservative   0; Mismatches   17; Indels   0; Gaps   0;

Qy   3 GGAGTTTCTTCATTCTGGTAGGTTCGTGGTCCTCTCGGCTTCAGGAATGAAGCTGTAGAA 62
              ||||||||||| ||||||||| ||||||||| |||||||||||| ||||||| ||
Db   403 GGAGTTTCTTCCTTCTGGTGGGTTCGTGGTCTGGTCGGCTTCAGGAGTGAAGCTGCAGAC 344

Qy   63 CTCTGCGATGCTGTTATGAACT 84
              ||| ||| || |||| ||
Db   343 CTTTGCGGTGAATGTTACAACT 322

RESULT 15
ABA64843/c
ID ABA64843 standard; DNA; 465 BP.
XX
AC ABA64843;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #13148.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.
XX
PS Claim 4; SEQ ID NO 13148; 639pp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The

CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note:The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 465 BP; 121 A; 129 C; 114 G; 101 T; 0 U; 0 Other;

Query Match   13.0%;   Score 54.8;   DB 4;   Length 465;
Best Local Similarity 79.3%;   Pred. No.7.6e-07;
Matches   65; Conservative   0; Mismatches 17; Indels   0; Gaps   0;

Qy   3 GGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAA 62
         ||||||||| ||||||||| ||||||||| |||||||||| |||||||||| |||
Db   403 GGAGTTTCTTCCTTCTGGTGGGTTCGTGTGGTCTCGGATGGCTTCAGGAGTGAAGCTGCGAGAC 344

Qy   63 CTCTGCGATGCTGTTATGAACT 84
         ||| |||||| ||| ||||
Db   343 CTTTGCGGGTGAAATGTTACAAACT 322

Search completed: April 25, 2004, 08:46:06
Job time : 114.862 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: April 24, 2004, 23:54:27 ; Search time 1108.58 Seconds
                                          (without alignments)
                                          16460.143 Million cell updates/sec

Title: US-10-051-835-16
Perfect score: 421
Sequence: 1 atggagttcttcattctgg.........ttctatgtcctttcagcgag 421

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
     1: gb_ba:*
     2: gb_htg:*
     3: gb_in:*
     4: gb_om:*
     5: gb_ov:*
     6: gb_pat:*
     7: gb_ph:*
     8: gb_pl:*
     9: gb_pr:*
    10: gb_ro:*
    11: gb_sts:*
    12: gb_sy:*
    13: gb_un:*
    14: gb_vi:*
    15: em_ba:*
    16: em_fun:*
    17: em_hum:*
    18: em_in:*
    19: em_mu:*
    20: em_om:*
    21: em_or:*
    22: em_ov:*
    23: em_pat:*
    24: em_ph:*
    25: em_pl:*
    26: em_ro:*
    27: em_sts:*
    28: em_un:*
    29: em_vi:*
    30: em_htg_hum:*
    31: em_htg_inv:*
    32: em_htg_other:*
    33: em_htg_mus:*
    34: em_htg_pln:*
    35: em_htg_rod:*
    36: em_htg_mam:*
    37: em_htg_vrt:*
    38: em_sy:*
    39: em_htgo_hum:*
    40: em_htgo_mus:*
    41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1	c	217.8	51.7	193829	9	AC012154	AC012154 Homo sapi
2	c	217.8	51.7	218508	2	AC020639	AC020639 Homo sapi
3	c	59.8	14.2	163497	9	AC021829	AC021829 Homo sapi
4		59.8	13.9	78031	9	AL392089	AL392089 Human DNA
5		58.2	13.8	143687	9	AC016572	AC016572 Homo sapi
6	c	58.2	13.8	146078	9	AC091874	AC091874 Homo sapi
7	c	57.8	13.7	174994	2	AC025306	AC025306 Homo sapi
8		57.2	13.6	145390	9	AC092963	AC092963 Homo sapi
9		57.2	13.6	158983	2	AC020598	AC020598 Homo sapi
10		57.2	13.6	178975	2	AC023222	AC023222 Homo sapi
11		57.2	13.6	179206	9	CNS01DS6	CNS01DS6 BAC seque
12		57.2	13.6	204158	9	AL133383	AL133383 Human DNA
13		57	13.5	75609	9	AC006456	AC006456 Homo sapi
14		57	13.5	172464	9	AC140889	AC140889 Homo sapi
15	c	57	13.5	174895	2	AC147084	AC147084 Pan trogl
16		56.6	13.4	2566	6	AX883871	AX883871 Sequence
17		56.6	13.4	2566	6	BD160580	BD160580 Primer fo
18		56.6	13.4	2566	9	AK024309	AK024309 Homo sapi
19		56.6	13.4	3687	9	BC041827	BC041827 Homo sapi
20	c	56.6	13.4	45303	9	AC092300	AC092300 Homo sapi
21	c	56.6	13.4	67605	9	AL441927	AL441927 Human DNA
22		56.6	13.4	99030	9	AC007038	AC007038 Homo sapi
23		56.6	13.4	104003	9	AC108064	AC108064 Homo sapi
24		56.6	13.4	110000	2	BX546444_08	Continuation (9 of
25		56.6	13.4	110227	9	AC121160	AC121160 Homo sapi
26		56.6	13.4	110227	2	AC125287	AC125287 Homo sapi
27	c	56.6	13.4	115168	2	AC084283	AC084283 Homo sapi
28		56.6	13.4	126619	9	AC092458	AC092458 Homo sapi
29	c	56.6	13.4	128000	2	AC004046	AC004046 Homo sapi
30		56.6	13.4	130726	9	AL365265	AL365265 Human DNA
31		56.6	13.4	130979	9	AL390719	AL390719 Human DNA
32	c	56.6	13.4	131060	9	HSDJ92C8	HSDJ92C8 Human DNA
33		56.6	13.4	147180	9	AC136469	AC136469 Homo sapi
34		56.6	13.4	149628	2	AC034128	AC034128 Homo sapi
35		56.6	13.4	149730	9	AL157387	AL157387 Human DNA
36	c	56.6	13.4	156667	9	AL136094	AL136094 Human DNA
37		56.6	13.4	156942	9	AC136470	AC136470 Homo sapi
38	c	56.6	13.4	157099	9	AC002451	AC002451 Homo sapi
39	c	56.6	13.4	166618	2	AC010323	AC010323 Homo sapi
40		56.6	13.4	174914	9	AC068874	AC068874 Homo sapi
41		56.6	13.4	181026	2	AL591642	AL591642 Homo sapi
42		56.6	13.4	183273	9	AC093898	AC093898 Homo sapi
43	c	56.6	13.4	197148	9	AC087651	AC087651 Homo sapi
44	c	56.6	13.4	197352	2	AC024725	AC024725 Homo sapi
45	c	56.6	13.4	210779	9	AC007450	AC007450 Homo sapi

ALIGNMENTS

RESULT 1
AC012154/c
LOCUS   AC012154   193829 bp   DNA   linear   PRI 28-JUN-2002
DEFINITION Homo sapiens 3 BAC RP11-48H24 (Roswell Park Cancer Institute Human
            BAC library) complete sequence.
ACCESSION   AC012154
VERSION   AC012154.16 GI:14578093
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 193829)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.

```
FEATURES             Location/Qualifiers
     source          1..193829
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-48H24"
     misc_feature    1..2008
                     /note="overlaps bases 152829..154828 of AC069070"
                     /function="clone overlap"
     repeat_region   244..365
                     /rpt_family="AluJb"
     repeat_region   366..412
                     /rpt_family="(CA)n"
     repeat_region   413..562
                     /rpt_family="AluJb"
     repeat_region   563..584
                     /rpt_family="AluJb"
     repeat_region   607..787
                     /rpt_family="(TAAA)n"
     repeat_region   complement(788..862)
                     /rpt_family="MER20"
     repeat_region   complement(901..1072)
                     /rpt_family="MER2"
     repeat_region   1684..1896
                     /rpt_family="MER2"
     repeat_region   complement(1897..2204)
                     /rpt_family="HAL1"
     repeat_region   2205..2228
                     /rpt_family="AluSg1"
     repeat_region   2229..2251
                     /rpt_family="HAL1"
     repeat_region   2252..2345
                     /rpt_family="(CA)n"
     repeat_region   2521..2646
                     /rpt_family="HAL1"
     repeat_region   3005..3126
                     /rpt_family="MLTI1"
```

Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,J., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newson,J., Newton,N., Nguyen,N., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
Rojas,A., Rojubokan,I., Rolfe,M., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
Gibbs,R.

TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 193829)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (21-OCT-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

REFERENCE   3 (bases 1 to 193829)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (07-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

REFERENCE   4 (bases 1 to 193829)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (10-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

REFERENCE   5 (bases 1 to 193829)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (28-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One .
            Baylor Plaza, Houston, TX 77030, USA

REFERENCE   6 (bases 1 to 193829)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (28-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

COMMENT   On Jun 30, 2001 this sequence version replaced gi:14547736.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

repeat_region	3181..3228	/rpt_family="HAL1"
repeat_region	3230..3273	/rpt_family="(TATG)n"
repeat_region	3503..3628	/rpt_family="AT_rich"
repeat_region	3629..4072	/rpt_family="LTR12"
repeat_region	4600..4699	/rpt_family="LTR12"
STS	complement(7298..7577)	/standard_name="76067"
repeat_region	8422..8458	/rpt_family="AluJo"
repeat_region	complement(8696..8873)	/rpt_family="AT_rich"
repeat_region	9120..9187	/rpt_family="MLT1D"
repeat_region	9695..10135	/rpt_family="(CATATA)n"
repeat_region	10323..10355	/rpt_family="MLT1C"
repeat_region	10355..10398	/rpt_family="(TG)n"
repeat_region	complement(10948..11063)	/rpt_family="(GA)n"
repeat_region	11038..11356	/rpt_family="L1P4"
repeat_region	12326..12366	/rpt_family="L1PA12"
repeat_region	13048..13086	/rpt_family="Harlequin"
repeat_region	13117..13360	/rpt_family="GA-rich"
repeat_region	14717..14744	/rpt_family="GA-rich"
repeat_region	14745..14778	/rpt_family="(TA)n"
repeat_region	14779..14822	/rpt_family="MADE1"
		/rpt_family="(TA)n"

```
Query Match            51.7%;   Score 217.8; DB 9;   Length 193829;
Best Local Similarity  94.9%;   Pred. No. 2.2e-54;
Matches  225; Conservative   0; Mismatches  12;   Indels  0;  Gaps  0;

Qy    185  TTTTTATGTGAGGACTCGGTCTCAGATATGGCTACTGGGATTCTGTTCAAACTGAGTGA 244
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33923  TTATTTTTATGCTAAATTTTTCAGATATGGCTACTGGGATTCTGTTCAAACTGAGTGA 33864

Qy    245  AGACATCATCATCTTCTTGAGATCTTCCTTGCTTTCTGGCCAAAGAGAGTTATTCCCTG 304
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33863  AGACATCATCATCTTCTTGAGATCTTCCTTGCTTTCTGGCCAAAGAGAGTTATTCCCTG 33804

Qy    305  ACCCGACTTTCCAAAGAGTACTAGTTCCTTTAGCAGTGATTAGAAACCTGATCTGAGTA 364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33803  ACCCGACTTTCCAAAGAGTACTAGTTCCTTTAGCAGTGATTAGAAACCTGATCTGAGTA 33744

Qy    365  CTACTGTGTGCTTGTGACTAGTGTGGTTTATTGTTTCTATGTCCTTTCAGCGAG 421
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33743  CTACTGTGTGTGCTTGTGACTAGTGTGGTTTATTGTTTCTATGTCCTTTCAGCGAG 33687
```

RESULT 2
ACC020639/c
LOCUS   AC020639   218508 bp   DNA   linear   HTG 27-JUN-2001
DEFINITION   Homo sapiens chromosome 3 clone RP11-32L10, WORKING DRAFT SEQUENCE,
             19 uncordered pieces.
ACCESSION   AC020639
VERSION   AC020639.18 GI:14547756
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens

             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 218508)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
             Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
             Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
             Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
             Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
             Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
             Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
             Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
             Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
             Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
             Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
             Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
             Garza,N., Gill,R., Gorrell,J.H., Guevara,N., Gunaratne,P., Hale,S.,
             Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
             Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
             Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
             Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
             Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
             Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
             Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
             Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
             Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
             Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
             Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
             Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newtson,N.,
             Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
             Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
             Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
             Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
             Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
             Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
             Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
             Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
             Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
             Wall,R., Wang,S., Ward-Moore,S., Warren,A., Wleczyk,R., Wooden,S.,
             Watlington,S., Williams,G., Williamson,A., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
             Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
             Weinstock,G. and Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 218508)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (07-JAN-2000) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jun 25, 2001 this sequence version replaced gi:14333746.
             Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
             --------------- Project Information ---------------
                Center project name: HMYW
                Center clone name: RP11-32L10
             --------------- Summary Statistics ---------------
                Sequencing vector: plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 91% of reads
                Chemistry: Dye-primer Bodipy: 9% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 230545 bases at least Q40
                Consensus quality: 245133 bases at least Q30
                Consensus quality: 253836 bases at least Q20
                Estimated insert size: 234517; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.9x in Q20 bases; sum-of-contigs estimation
             ----
             * NOTE: Estimated insert size may differ from sequence length

```
Db      165309 CTACTGTGTGTGCTTGTGACTAGTGTGGTTTATTGTTTCATGTCCTTTCAGCGAG 165253

RESULT 3
LOCUS       AC021829/c           163497 bp    DNA     linear   PRI 23-JAN-2002
DEFINITION  Homo sapiens chromosome 11, clone RP11-436H16, complete sequence.
ACCESSION   AC021829
VERSION     AC021829.8  GI:17647005
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163497)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-436H16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163497)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,G., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,C., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,J., McCurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 163497)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,G.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., Lakocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McKernan,K., McPheeters,M., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,N., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT   On Dec 13, 2001 this sequence version replaced gi:17426390.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
```

```
*  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
*  consists of 19 contigs. The true order of the pieces
*  is not known and their order in this sequence record is
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.
*        1    70229: contig of 70229 bp in length
*    70230    70329: gap of unknown length
*    70330   103341: contig of 33012 bp in length
*   103342   103441: gap of unknown length
*   103442   136164: contig of 32723 bp in length
*   136165   136264: gap of unknown length
*   136265   155333: contig of 19069 bp in length
*   155334   155433: gap of unknown length
*   155434   160564: contig of 5131 bp in length
*   160565   160664: gap of unknown length
*   160665   167685: contig of 7021 bp in length
*   167686   167785: gap of unknown length
*   167786   174526: contig of 6741 bp in length
*   174527   174626: gap of unknown length
*   174627   178013: contig of 3387 bp in length
*   178014   178113: gap of unknown length
*   178114   183160: contig of 5047 bp in length
*   183161   183260: gap of unknown length
*   183261   186615: contig of 3355 bp in length
*   186616   186715: gap of unknown length
*   186716   190516: contig of 3801 bp in length
*   190517   190616: gap of unknown length
*   190617   194999: contig of 4383 bp in length
*   195000   195099: gap of unknown length
*   195100   199336: contig of 4237 bp in length
*   199337   199436: gap of unknown length
*   199437   203453: contig of 4017 bp in length
*   203454   203553: gap of unknown length
*   203554   206727: contig of 3174 bp in length
*   206728   206827: gap of unknown length
*   206828   210654: contig of 3827 bp in length
*   210655   210754: gap of unknown length
*   210755   213431: contig of 2677 bp in length
*   213432   213531: gap of unknown length
*   213532   215931: contig of 2400 bp in length
*   215932   216031: gap of unknown length
*   216032   218508: contig of 2477 bp in length.
FEATURES             Location/Qualifiers
     source          1..218508
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-32L10"
ORIGIN

Query Match       51.7%;  Score 217.8;  DB 2;  Length 218508;
Best Local Similarity 94.9%;  Pred. No. 2.2e-54;
Matches  225;  Conservative  0;  Mismatches  12;  Indels  0;  Gaps  0;

Qy   185 TTTTTATGTGAGGAGCGGTCTCAGATATGGCTACTGGGATTCGTTCAAACTGAGTGA 244
             | || |  |||||||||||||||||||||||||||||||||||||||||||||
Db 165489 TTATTTTATGCTAAATTTTTTCCAGATATGGCTACTGGGATTCGTTCAAACTGAGTGA 165430

Qy   245 AGACATCATCATCTTCTTTGAGATCTTCCTTGCTTCTCGGCCAAAGAAGTTATTCCCTG 304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165429 AGACATCATCATCTTCTTTGAGATCTTCCTTGCTTCTCGGCCAAAGAAGTTATTCCCTG 165370

Qy   305 ACCCGACTTTCCAAAGAGTACTAGTCCTTTAGCAGTGATTAGAAACCTGATCTGAGTA 364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165369 ACCCGACTTTCCAAAGAGTACTAGTCCTTTAGCAGTGATTAGAAACCTGATCTGAGTA 165310

Qy   365 CTACTGTGTGTGCTTGTGACTAGTGTGGTTTATTGTTTCATGTCCTTTCAGCGAG 421
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
         Contact: sequence submissions@genome.wi.mit.edu
         ------- Project Information
         Center project name: L5970
         Center clone name: 436_H_16

FEATURES         Location/Qualifiers
source           1..163497
                 /organism="Homo sapiens"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:9606"
                 /chromosome="11"
                 /map="11"
                 /clone="RP11-436H16"
                 /clone_lib="RPCI-11 Human Male BAC"
repeat_region    complement(1..2132)
                 /rpt_family="L1M1"
repeat_region    complement(2386..2671)
                 /rpt_family="AluSg"
repeat_region    complement(3787..3926)
                 /rpt_family="MIR"
repeat_region    4136..4174
                 /rpt_family="(CA)n"
repeat_region    complement(5803..6089)
                 /rpt_family="AluSg"
repeat_region    complement(6390..6891)
                 /rpt_family="MLT1D"
repeat_region    7240..7288
                 /rpt_family="AT_rich"
repeat_region    7569..7733
                 /rpt_family="L1ME"
repeat_region    7734..7824
                 /rpt_family="(TA)n"
repeat_region    7959..8092
                 /rpt_family="L1ME"
repeat_region    8140..8356
                 /rpt_family="L1ME"
repeat_region    8363..8520
                 /rpt_family="L1PA16"
repeat_region    8696..8886
                 /rpt_family="L1ME"
repeat_region    10445..10530
                 /rpt_family="(GAAA)n"
unsure           10782..10787
                 /note="<30 qual SNGL region"
repeat_region    11039..11078
                 /rpt_family="(T)n"
repeat_region    complement(11094..11279)
                 /rpt_family="MLT1F"
repeat_region    complement(11403..11483)
                 /rpt_family="MLT1F"
repeat_region    11548..11588
                 /rpt_family="L1ME"
repeat_region    complement(12712..12893)
                 /rpt_family="(CA)n"
repeat_region    13048..13478
                 /rpt_family="MIR"
repeat_region    13478..14753
                 /rpt_family="MER68"
repeat_region    14702..14753
                 /rpt_family="(TC)n"
repeat_region    15447..15719
                 /rpt_family="L3"
repeat_region    15887..16080
                 /rpt_family="L3"
repeat_region    16098..16478
                 /rpt_family="L1MD3"
repeat_region    16494..17646
                 /rpt_family="L1MD3"
repeat_region    complement(17647..17945)
                 /rpt_family="AluY"
repeat_region    17946..18682
                 /rpt_family="L1MD3"
repeat_region    complement(19872..20357)
                 /rpt_family="L1M4c"
unsure           20358..20363
                 /note="<30 qual SNGL region"
repeat_region    20404..21580
                 /rpt_family="LTR12"
repeat_region    complement(21583..22973)
                 /rpt_family="L1M4c"
repeat_region    complement(23009..23754)
                 /rpt_family="MER97C"
repeat_region    complement(24572..24870)
                 /rpt_family="MER58B"
repeat_region    25237..25275
                 /rpt_family="(CA)n"
repeat_region    25345..25720
                 /rpt_family="THE1C"
repeat_region    27806..27857
                 /rpt_family="AT_rich"
repeat_region    27915..28123
                 /rpt_family="MER30"
repeat_region    28522..28555
                 /rpt_family="A-rich"
repeat_region    complement(28779..30123)
                 /rpt_family="L1PA9"
repeat_region    complement(30172..31702)
                 /rpt_family="L1M4"
repeat_region    complement(31707..31783)
                 /rpt_family="L1M4"
repeat_region    complement(31838..32080)
                 /rpt_family="L1M4"
repeat_region    complement(32100..32281)
                 /rpt_family="L1M4"
repeat_region    32753..32782
                 /rpt_family="AT_rich"
repeat_region    33036..33191
                 /rpt_family="AT_rich"
repeat_region    33313..33562
                 /rpt_family="MIR"
repeat_region    33656..33949
                 /rpt_family="AluSg"
repeat_region    complement(34991..35652)
                 /rpt_family="L1PA14"
repeat_region    35653..35676
                 /rpt_family="(TTTG)n"
repeat_region    complement(35677..35954)
                 /rpt_family="AluSg"
repeat_region    complement(35955..36138)
                 /rpt_family="L1PA14"
repeat_region    36142..36597
                 /rpt_family="L1P"
repeat_region    36910..36932
                 /rpt_family="AT_rich"
repeat_region    37002..37030
                 /rpt_family="AT_rich"
repeat_region    37940..39330
                 /rpt_family="pTR5"
repeat_region    39419..39802
                 /rpt_family="MER67C"
repeat_region    40122..40151
                 /rpt_family="AT_rich"
repeat_region    complement(40155..40512)
                 /rpt_family="MER114"
```

```
Query Match        14.2%;    Score 59.8;   DB 9;    Length 163497;
Best Local Similarity 90.1%;   Pred. No. 6.7e-07;   Indels 0;   Gaps 0;
Matches  64;   Conservative  0;   Mismatches  7;   Indels  0;

Qy      2 TGGAGTTCTTCATTCTGGTAGGTTCGTGGCTTCTCTGGCTTCAGGAATGAAGCTGTAGA 61
          ||||||||||| |||||||| |||||| ||| ||| |||| |||||||| ||||||| ||||
Db  39200 TGGAGTTCTTCCTTCTGGTGGGTTCATGGTCTCGCTTCGCTGGCTTCAGGAGTGAAGCTGCAGA 39141

Qy     62 ACTCTGCGATG 72
          |||||||||||
Db  39140 CCTCTGCGATG 39130
```

RESULT 5
AL016572
LOCUS   AC016572   143687 bp   DNA   linear   PRI 17-OCT-2001
DEFINITION   Homo sapiens chromosome 5 clone CTB-22L19, complete sequence.
ACCESSION   AC016572
VERSION   AC016572.6 GI:16195210
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 143687)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 143687)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
               Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 143687)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (01-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
               Drive, Walnut Creek, CA 94598, USA
REFERENCE   4 (bases 1 to 143687)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE   Direct Submission
JOURNAL   Submitted (17-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
               Drive, Walnut Creek, CA 94598, USA
COMMENT   On Oct 17, 2001 this sequence version replaced gi:8886979.
               Draft Sequence Produced by DOE Joint Genome Institute
               www.jgi.doe.gov
               Finishing Completed at Stanford Human Genome Center
               www-shgc.stanford.edu
               Quality: Phrap Quality >40 99.8% of Sequence;
               Estimated Total Number of Errors is 0.2.
               STS Content:
               WI-15997 G21714.
FEATURES   Location/Qualifiers
     source   1..143687
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-22L19"
ORIGIN
Query Match   13.8%; Score 58.2; DB 9; Length 143687;
Best Local Similarity 88.7%; Pred. No. 2e-06;
Matches   63; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

Qy   2 TGGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTGGCTTCAGGAATGAAGCTGTAGA 61
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   133912 TGGAGTTCTTCCTTCTCGGTGGGTTCGTGGGTCTCTGGCTTCAGGAGTGAAGCTGCAGA 133971

Qy   62 ACTCTGCGATG 72
              |||||||||||
Db   133972 CCTTAGAGATG 133982

RESULT 6
AC091874/c
LOCUS   AC091874   146078 bp   DNA   linear   PRI 03-APR-2003
DEFINITION   Homo sapiens chromosome 5 clone CTD-2593A12, complete sequence.
ACCESSION   AC091874
VERSION   AC091874.4 GI:29501850
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 146078)

RESULT 4
AL392089
LOCUS   AL392089   78031 bp   DNA   linear   PRI 21-DEC-2000
DEFINITION   Human DNA sequence from clone RP11-31F19 on chromosome 9p24.1-24.3,
               complete sequence.
ACCESSION   AL392089
VERSION   AL392089.7 GI:11967571
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 78031)
AUTHORS   Lloyd,D.
TITLE   Direct Submission
JOURNAL   Submitted (20-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
               CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
               requests: clonerequest@sanger.ac.uk
COMMENT   On Dec 22, 2000 this sequence version replaced gi:11691527.
               During sequence assembly data is compared from overlapping clones.
               Where differences are found these are annotated as variations
               together with a note of the overlapping clone name. Note that the
               variation annotation may not be found in the sequence submission
               corresponding to the overlapping clone, as we submit sequences with
               only a small overlap as described above.
               This sequence has been finished according to sequence map criteria
               as follows. An attempt is made to resolve all sequencing problems,
               such as compressions and repeats, but not necessarily within known
               annotated repeat sequence elements. Where the sequence is
               ambiguous, there is an annotation using the 'unsure' feature key.
               The following abbreviations are used to associate primary accession
               numbers given in the feature table with their source databases:
               Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
               on the WORMPEP database can be found at
               http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
               was generated from part of bacterial clone contigs of human
               chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
               Group. Further information can be found at
               http://www.sanger.ac.uk/HGP/Chr9
               RP11-31F19 is from the library RPCI-11.1 constructed by the group
               of Pieter de Jong. For further details see
               http://www.chori.org/bacpac/home.htm
               VECTOR: pBACe3.6
               IMPORTANT: This sequence is not the entire insert of clone
               RP11-31F19 It may be shorter because we sequence overlapping
               sections only once, except for a 100 base overlap.
               The true left end of clone RP11-130C19 is at 77992 in this
               sequence. The true right end of clone RP11-165F24 is at 100 in this
               sequence.
FEATURES   Location/Qualifiers
     source   1..78031
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="p24.1-24.3"
                     /clone="RP11-31F19"
                     /clone_lib="RPCI-11.1"
ORIGIN
Query Match   13.9%; Score 58.6; DB 9; Length 78031;
Best Local Similarity 82.7%; Pred. No. 1.6e-06;
Matches   67; Conservative   0; Mismatches   14; Indels   0; Gaps   0;

Qy   4 GAGGTTCTTCATTCTGGTAGGTTCGTGGTCTCTGGCTTCAGGAATGAAGCTGTAGAAC 63
              ||||| |||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   32881 GAGGTTCTTCCTTCTCGGTGGGTTCGTGGGTCTCTGGCTTCACTGGCTTCAGGAGTGAAGCTGCAGACC 32940

Qy   64 TCTCGGATGCTGTTATGAACT 84
              |||||||||||||| ||||||
Db   32941 TTCGCGGTGAGTGTAACAGCT 32961

```
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 146078)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 146078)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (22-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 146078)
AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE       Direct Submission
JOURNAL     Submitted (03-APR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Apr 3, 2003 this sequence version replaced gi:29150342.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
            NOTE: Transposon sequencing failed to verify the number of repeat
            copies 12696-13210.  Unsure number of repeat copies 12696-13210.
FEATURES             Location/Qualifiers
     source          1..146078
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2593A12"
     misc_feature    12696..13210
                     /note="NOTE:  Transposon sequencing failed to verify the
                     number of repeat copies 12696-13210.  Unsure number of
                     repeat copies 12696-13210."
ORIGIN

Query Match         13.8%;  Score 58.2;  DB 9;  Length 146078;
Best Local Similarity   88.7%;  Pred. No. 2e-06;
Matches   63;  Conservative  0;  Mismatches  8;  Indels  0;  Gaps  0;

Qy           2  TGGAGTTCTCATTCTGGTAGGTTCGTGGTCCTCTGGCTTCAGGAATGAAGCTGTAGA 61
                ||||||||||||||||||||||  |||||||||||||||||||||||||||||||
Db       90111  TGGAGTTCTCTTCCTCTTCGGTGGGTTCGTGGGTCCTCTGGCTTCAGGAGTGAAGCTGCAGA 90052

Qy          62  ACTCTGCGATG 72
                ||||||||||
Db       90051  CCTTAGAGATG 90041
```

```
RESULT 7
AC025306/c
LOCUS       AC025306             174994 bp    DNA     linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens clone RP11-24F7, WORKING DRAFT SEQUENCE, 20 unordered
            pieces.
ACCESSION   AC025306
VERSION     AC025306.3  GI:8072565
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174994)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Direct Submission
JOURNAL     Submitted (03-APR-2003) this sequence version replaced
```

```
TITLE
JOURNAL

COMMENT
```

```
AUTHORS     Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7328744.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
            Center project name: L449
            Center clone name: 24 F7
            ---------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 164074 bases at least Q40
            Consensus quality: 169122 bases at least Q30
            Consensus quality: 171310 bases at least Q20
            Insert size: 172000; agarose-fp
            Insert size: 173094; sum-of-contigs
            Quality coverage: 4.4 in Q20 bases; agarose-fp
            Quality coverage: 4.4 in Q20 bases; sum-of-contigs
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*            1     2060: contig of 2060 bp in length
*         2061     2160: gap of 100 bp
*         2161     3367: contig of 1207 bp in length
*         3368     3467: gap of 100 bp
*         3468     5552: contig of 2085 bp in length
*         5553     5652: gap of 100 bp
*         5653     7724: contig of 2072 bp in length
*         7725     7824: gap of 100 bp
*         7825    11645: contig of 3821 bp in length
*        11646    11745: gap of 100 bp
*        11746    16959: contig of 5214 bp in length
*        16960    17059: gap of 100 bp
*        17060    21538: contig of 4479 bp in length
*        21539    21638: gap of 100 bp
*        21639    27069: contig of 5431 bp in length
*        27070    27169: gap of 100 bp
*        27170    33103: contig of 5934 bp in length
*        33104    33203: gap of 100 bp
*        33204    39795: contig of 6592 bp in length
*        39796    39895: gap of 100 bp
```

```
REFERENCE   2  (bases 1 to 174994)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
```

Best Local Similarity 89.9%; Pred. No. 2.7e-06;
Matches 62; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy       4 GAGTTCTTCATTCTGGTAGGTTCGTGGTCTCTGGCTTCAGGAATGAAGCTGTAGAAC  63
           |||||||||||||||||||||||||||| ||||||||||||| |||||||||||||||
Db  105787 GAGTTCTTCATTCTGGTAGGTTCGTGGTCTCACTGGCTTCAGGAGTGAAGCTGCAGACC 105728

Qy      64 TCTGCGATG  72
           |||||||||
Db  105727 TTCGCGGTG 105719
```

RESULT 8
ACCESSION   AC092963
LOCUS   AC092963   145390 bp   DNA   linear   PRI 30-SEP-2002
DEFINITION   Homo sapiens 3q BAC RP11-396A13 (Roswell Park Cancer Institute
             Human BAC Library) complete sequence.
ACCESSION   AC092963
VERSION   AC092963.4 GI:18958598
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 145390)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Alebrooks,S.L., Amaratunge.H.C., Are,J.R., Ayele,M., Banks,T.,
             Barbaria,J., Benton,J., Bimage,K., Blankenship,K., Bonnin,D.,
             Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
             Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
             Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
             Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
             Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
             Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
             Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.
             Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
             Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
             Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
             Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
             Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
             Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
             Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
             Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
             Jackson,L.E., Jacobson,B., Jia,Y., Johnson,J., Jolivet,S.,
             Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,C., Korvah,J.,
             Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
             Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
             Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
             Ma,J., Maheshwari,M., Mapua,P., Martin,A., Martindale,A.,
             Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
             Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
             Moore,C., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
             Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
             Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
             Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
             Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
             Rojas,G., Rojubokan,I., Rolfe,M., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
             Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
             Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
             Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
             Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
             Vera,V., Villalon,D., vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
             Warren,R., Washington,C., Watlington,S., Williams,G.,
             Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
             Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
             Gibbs,R.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 145390)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
*    39896   46583: contig of 6688 bp in length
*    46584   46683: gap of 100 bp
*    46684   55012: contig of 8329 bp in length
*    55013   55112: gap of 100 bp
*    55113   63297: contig of 8185 bp in length
*    63298   63397: gap of 100 bp
*    63398   74634: contig of 11237 bp in length
*    74635   74734: gap of 100 bp
*    74735   84510: contig of 9776 bp in length
*    84511   84610: gap of 100 bp
*    84611   95119: contig of 10509 bp in length
*    95120   95219: gap of 100 bp
*    95220  106937: contig of 11718 bp in length
*   106938  107037: gap of 100 bp
*   107038  122346: contig of 15309 bp in length
*   122347  122446: gap of 100 bp
*   122447  145374: contig of 22928 bp in length
*   145375  145474: gap of 100 bp
*   145475  174994: contig of 29520 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1..174994
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-24F7"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..2060
                     /note="assembly_fragment"
     misc_feature    2161..3367
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    3468..5552
                     /note="assembly_fragment"
     misc_feature    5653..7724
                     /note="assembly_fragment"
     misc_feature    7825..11645
                     /note="assembly_fragment"
     misc_feature    11746..16959
                     /note="assembly_fragment"
     misc_feature    17060..21538
                     /note="assembly_fragment"
     misc_feature    21639..27069
                     /note="assembly_fragment"
     misc_feature    27170..33103
                     /note="assembly_fragment"
     misc_feature    33204..39795
                     /note="assembly_fragment"
     misc_feature    39896..46583
                     /note="assembly_fragment"
     misc_feature    46684..55012
                     /note="assembly_fragment"
     misc_feature    55113..63297
                     /note="assembly_fragment"
     misc_feature    63398..74634
                     /note="assembly_fragment"
     misc_feature    74735..84510
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    84611..95119
                     /note="assembly_fragment"
     misc_feature    95220..106937
                     /note="assembly_fragment"
     misc_feature    107038..122346
                     /note="assembly_fragment"
     misc_feature    122447..145374
                     /note="assembly_fragment"
     misc_feature    145475..174994
                     /note="assembly_fragment"
ORIGIN
Query Match  13.7%;  Score 57.8;  DB 2;  Length 174994;
```

REFERENCE   3 (bases 1 to 145390)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (27-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4 (bases 1 to 145390)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5 (bases 1 to 145390)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (30-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Feb 27, 2002 this sequence version replaced gi:17921172.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
            STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
            Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
            Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

FEATURES   Location/Qualifiers
     source   1..145390
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-396A13"
     misc_feature   4..1904
                     /note="overlaps bases 69098..70998 of clone AC092970"
                     /function="clone overlap"
     repeat_region   complement(7..619)
                     /rpt_family="L1M1"
     repeat_region   656..853
                     /rpt_family="L1M1"
     repeat_region   854..885
                     /rpt_family="AluSq"

     repeat_region   /rpt_family="(TAAA)n"
                     complement(890..1249)
     repeat_region   /rpt_family="THE1C"
                     complement(1252..2594)
     repeat_region   /rpt_family="L1M1"
                     complement(2595..2896)
     repeat_region   /rpt_family="AluSg"
                     complement(2897..2928)
     repeat_region   /rpt_family="L1M1"
                     complement(2929..3220)
     repeat_region   /rpt_family="AluJo"
                     complement(3221..3973)
     repeat_region   /rpt_family="L1M1"
                     3975..4011
     repeat_region   /rpt_family="(TG)n"
                     complement(4012..4231)
     repeat_region   /rpt_family="L1M2"
                     4229..4346
     repeat_region   /rpt_family="L1M2"
                     complement(4434..4973)
     repeat_region   /rpt_family="L1M2"
                     complement(5299..5353)
     repeat_region   /rpt_family="MIR"
                     5354..5396
     repeat_region   /rpt_family="(TG)n"
                     complement(5397..5524)
     repeat_region   /rpt_family="MIR"
                     5940..6310
     repeat_region   /rpt_family="L1M4"
                     complement(6318..6934)
     repeat_region   /rpt_family="MER41B"
                     complement(6964..7398)
     repeat_region   /rpt_family="L1PA7"
                     7418..7521
     repeat_region   /rpt_family="L1M4"
                     7753..8085
     repeat_region   /rpt_family="L1M4"
                     8243..8430
     repeat_region   /rpt_family="L1ME3A"
                     8458..8490
     repeat_region   /rpt_family="(TTTA)n"
                     complement(8491..8576)
     repeat_region   /rpt_family="AluSc"
                     8590..9363
     repeat_region   /rpt_family="L1ME3A"
                     complement(9380..9932)
     repeat_region   /rpt_family="MLT2A"
                     10181..10208
     repeat_region   /rpt_family="AT_rich"
                     13160..13203
     repeat_region   /rpt_family="(TTTG)n"
                     complement(13204..13490)
     repeat_region   /rpt_family="AluSg"
                     complement(14232..14383)
     repeat_region   /rpt_family="MER92A"
                     complement(14405..15301)
     repeat_region   /rpt_family="L1MD1"
                     complement(15302..15691)
     repeat_region   /rpt_family="MLT1B"
                     15715..15790
     repeat_region   /rpt_family="MLT1A1"
                     15791..16081
     repeat_region   /rpt_family="AluSx"
                     16082..16322
     repeat_region   /rpt_family="MLT1A1"
                     complement(16327..16465)
     repeat_region   /rpt_family="L1MD1"
                     complement(16445..16817)

Query Match
Best Local Similarity   13.6%;   Score 57.2; DB 9; Length 145390;
Matches 62; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
                        88.6%; Pred. No. 4.1e-06;

```
Qy          3 GGAGTTCTTCATTCGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAA 62
              |||||||||||| ||||||| |||||||| ||||||||||| ||||||||| |||
Db      20401 GGAGTTCTTCTTCCTTCGTGGGTCGGGTTCGTGGTCTCACTGGCTTCAGGAGTGAAGCTGCAGAC 20460

Qy         63 CTCTGCGATG 72
              || |||||||
Db      20461 CTTTGCGGTG 20470

RESULT 9
AC020598
LOCUS       AC020598              158983 bp    DNA     linear   HTG 18-FEB-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-404J16, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC020598
VERSION     AC020598.6  GI:12963047
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158983)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 158983)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (05-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Feb 18, 2001 this sequence version replaced gi:9858987.
            ----------------------- Genome Center -----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            --------------- Project Information ---------------
            Center project name: H_NH0404J16
            ---------------- Summary Statistics ----------------
            Sequencing vector: M13; 85%
            Sequencing vector: plasmid; 15%
            Chemistry: Dye-primer ET; 85% of reads
            Chemistry: Dye-terminator Big Dye; 15% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 153629 bases at least Q40
            Consensus quality: 155139 bases at least Q30
            Consensus quality: 155860 bases at least Q20
            Insert size: 145000; agarose-fp
            Insert size: 157883; sum-of-contigs
            Quality coverage: 6.88 in Q20 bases; agarose-fp
            Quality coverage: 6.34 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *           1    2675: contig of 2675 bp in length
            *        2676    2775: gap of unknown length
            *        2776    7991: contig of 5216 bp in length
            *        7992    8091: gap of unknown length
            *        8092   14234: contig of 6143 bp in length
            *       14235   14334: gap of unknown length
            *       14335   28444: contig of 14110 bp in length
            *       28445   28544: gap of unknown length
            *       28545   49191: contig of 20647 bp in length
            *       49192   49291: gap of unknown length
            *       49292   69715: contig of 20424 bp in length
            *       69716   69815: gap of unknown length
```

```
            *       69816  109840: contig of 40025 bp in length
            *      109841  109940: gap of unknown length
            *      109941  150702: contig of 40762 bp in length
            *      150703  150802: gap of unknown length
            *      150803  152468: contig of 1666 bp in length
            *      152469  152568: gap of unknown length
            *      152569  154401: contig of 1833 bp in length
            *      154402  154501: gap of unknown length
            *      154502  156475: contig of 1974 bp in length
            *      156476  156575: gap of unknown length
            *      156576  158983: contig of 2408 bp in length.
FEATURES             Location/Qualifiers
     source          1..158983
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-404J16"
     misc_feature    1..2675
                     /note="assembly_name:Contig10"
     misc_feature    2776..7991
                     /note="assembly_name:Contig11"
     misc_feature    8092..14234
                     /note="assembly_name:Contig12"
     misc_feature    14335..28444
                     /note="assembly_name:Contig13"
     misc_feature    28545..49191
                     /note="assembly_name:Contig14
                     clone_end:T7
                     vector_side:right"
     misc_feature    49292..69715
                     /note="assembly_name:Contig15"
     misc_feature    69816..109840
                     /note="assembly_name:Contig16"
     misc_feature    109941..150702
                     /note="assembly_name:Contig17
                     clone_end:SP6
                     vector_side:right"
     misc_feature    150803..152468
                     /note="assembly_name:Contig6"
     misc_feature    152569..154401
                     /note="assembly_name:Contig7"
     misc_feature    154502..156475
                     /note="assembly_name:Contig8"
     misc_feature    156576..158983
                     /note="assembly_name:Contig9"
ORIGIN

Query Match             13.6%;  Score 57.2;  DB 2;  Length 158983;
Best Local Similarity   88.6%;  Pred. No. 4e-06;
Matches   62;  Conservative   0;  Mismatches   8;  Indels   0;  Gaps   0;

Qy          3 GGAGTTCTTCATTCGGTAGGTTCGTGGTCTCTCGGCTTCAGGAATGAAGCTGTAGAA 62
              |||||||||||| ||||||| |||||||| ||||||||||| ||||||||| |||
Db      82864 GGAGTTCTTCTTCCTTCGTGGGTCGGGTTCGTGGTCTCACTGGCTTCAGGAGTGAAGCTGCAGAC 82923

Qy         63 CTCTGCGATG 72
              || |||||||
Db      82924 CTTTGCGGTG 82933

RESULT 10
AC023222
LOCUS       AC023222              178975 bp    DNA     linear   HTG 17-MAY-2000
DEFINITION  Homo sapiens chromosome 7 clone RP11-396A13 map 7, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC023222
VERSION     AC023222.3  GI:7884386
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 178975)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 7, clone RP11-396A13
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 178975)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Penestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 17, 2000 this sequence version replaced gi:7622394.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
            Center project name: L6676
            Center clone name: 396_A13
            ---------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 171366 bases at least Q40
            Consensus quality: 175529 bases at least Q30
            Consensus quality: 176962 bases at least Q20
            Insert size: 192000; agarose-fp
            Insert size: 177875; sum-of-contigs
            Quality coverage: 4.4 in Q20 bases; agarose-fp
            Quality coverage: 4.8 in Q20 bases; sum-of-contigs
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 12 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *      1      1862: contig of 1862 bp in length
          *   1863      1962: gap of 100 bp
          *   1963      5596: contig of 3634 bp in length
          *   5597      5696: gap of 100 bp
          *   5697     12695: contig of 6999 bp in length
          *  12696     12795: gap of 100 bp
          *  12796     24148: contig of 11353 bp in length
          *  24149     24248: gap of 100 bp
          *  24249     34426: contig of 10178 bp in length
          *  34427     34526: gap of 100 bp
          *  34527     44838: contig of 10312 bp in length
          *  44839     44938: gap of 100 bp
          *  44939     58380: contig of 13442 bp in length
          *  58381     58480: gap of 100 bp
          *  58481     79218: contig of 20738 bp in length
          *  79219     79318: gap of 100 bp
          *  79319     97949: contig of 18631 bp in length
          *  97950     98049: gap of 100 bp
          *  98050    117961: contig of 19912 bp in length
          * 117962    118061: gap of 100 bp
          * 118062    144166: contig of 26105 bp in length
          * 144167    144266: gap of 100 bp
          * 144267    178975: contig of 34709 bp in length.
FEATURES             Location/Qualifiers
     source          1..178975
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-396A13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1862
                     /note="assembly_fragment"
     misc_feature    1963..5596
                     /note="assembly_fragment"
     misc_feature    5697..12695
                     /note="assembly_fragment"
     misc_feature    12795..24148
                     /note="assembly_fragment"
     misc_feature    24244..34426
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    34527..44838
                     /note="assembly_fragment"
     misc_feature    44939..58380
                     /note="assembly_fragment"
     misc_feature    58481..79218
                     /note="assembly_fragment"
     misc_feature    79319..97949
                     /note="assembly_fragment"
     misc_feature    98050..117961
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    118062..144166
                     /note="assembly_fragment"
     misc_feature    144267..178975
                     /note="assembly_fragment"
ORIGIN

Query Match             13.6%;  Score 57.2;  DB 2;  Length 178975;
Best Local Similarity   88.6%;  Pred. No. 4e-06;
Matches   62;  Conservative   0;  Mismatches   8;  Indels   0;  Gaps   0;

Qy      3 GGAGTTCTTCATTCTCGGTAGGTTCGTGGTCTCCTCGGCTTCAGGAATGAAGCTGTAGAA 62
          ||||||||||  ||||||| |||||||||||| |||||||| ||||||||||||||||| |
Db   6096 GGAGTTCTTCCTTCTCGGTGGTGGGTTCGTGGTTCGGGCTTCACTGGCTTCAGGAGTGAAGCTGCGCAGAC 6155

Qy     63 CTCTGCGATG 72
          || |||||||
Db   6156 CTTCGCGATG 6165

RESULT 11
CNS01DS6
LOCUS       CNS01DS6       179206 bp    DNA    linear    PRI 18-APR-2002
DEFINITION  BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 367K01
            of library CITB_978_SKB from chromosome 2 of Homo sapiens (Human).
ACCESSION   AL121656
VERSION     AL121656.2  GI:7159617
KEYWORDS    SPG4 genomic DNA interval.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 179206)
AUTHORS     Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,
```

```
  TITLE       Artiguenave,R., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,
              Brottier,P., Catolico,L., Barbe,V., Burgunder,J.M.,
              Prud'Homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
              Weissenbach,J.
  TITLE       Spastin, a novel AAA protein, is altered in the most frequent form
              of autosomal dominant spastic paraplegia
  JOURNAL     Nat. Genet. (1999) In press
  REFERENCE   2  (bases 1 to 179206)
  AUTHORS     Genoscope.
  TITLE       Direct Submission
  JOURNAL     Submitted (18-APR-2002) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
  COMMENT     On Mar 6, 2000 this sequence version replaced gi:6981752.
  FEATURES             Location/Qualifiers
     source            1..179206
                       /organism="Homo sapiens"
                       /mol_type="genomic DNA"
                       /db_xref="taxon:9606"
                       /chromosome="2"
                       /clone="367K01"
                       /clone_lib="CITB_978_SKB"
  ORIGIN
  Query Match          13.6%;    Score 57.2;  DB 9;   Length 179206;
  Best Local Similarity 88.6%;   Pred. No. 4e-06;
  Matches    62;  Conservative    0;  Mismatches   8;  Indels    0;  Gaps     0;

  Qy        3 GGAGTTCTTCATTCTGGTAGGTTCGTGGTCTCCTGGCTTCAGGAATGAAGCTGTAGAA 62
              |||||||||||||||||||||||||||||||| |||||||||||||||||||| |||||
  Db   107000 GGAGTTCTTCTTCCTTCTGGTGGGTTCGTGGGTTCGTGGTCTCACTGGCTTCAGGAGTGAAGCTGCAGAC 107059

  Qy       63 CTCTGCGATG 72
              ||||||||||
  Db   107060 CTTTGCGGTG 107069

  RESULT 12
  AL133383
  LOCUS       AL133383             204158 bp    DNA       linear   PRI 04-APR-2001
  DEFINITION  Human DNA sequence from clone GS1-204I12 on chromosome 1. Contains
              ESTs, STSs and GSSs. Contains a novel gene, complete sequence.
  ACCESSION   AL133383
  VERSION     AL133383.10  GI:10119658
  KEYWORDS    HTG.
  SOURCE      Homo sapiens (human)
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE   1  (bases 1 to 204158)
  AUTHORS     Hall,R.
  TITLE       Direct Submission
  JOURNAL     Submitted (05-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
              requests: clonerequest@sanger.ac.uk
  COMMENT     On Sep 12, 2000 this sequence version replaced gi:10086005.
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              The following abbreviations are used to associate primary accession
              numbers given in the feature table with their source databases:
              Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
              on the WORMPEP database can be found at
              http://www.sanger.ac.uk/Projects/C_elegans/wormpep/wormpep GS1-204I12 is
              from the library Genome_Systems_Release1 VECTOR: pBeloBAC11
              This sequence is the entire insert of clone GS1-204I12 This
              sequence has been finished according to sequence map criteria as
              follows. An attempt is made to resolve all sequencing problems,
              such as compressions and repeats, but not necessarily within known
              annotated repeat sequence elements.  Where the sequence is
```

```
              ambiguous, there is an annotation using the 'unsure' feature key.
              This sequence was generated by the Sanger Centre from part of a
              human chromosome 1 bacterial clone contig constructed by John
              Carpten et al, NHGRI, NIH. Further information can be found at
              http://www.sanger.ac.uk/HGP/Chr1.
  FEATURES             Location/Qualifiers
     source            1..204158
                       /organism="Homo sapiens"
                       /mol_type="genomic DNA"
                       /db_xref="taxon:9606"
                       /chromosome="1"
                       /clone="GS1-204I12"
                       /clone_lib="Genome_Systems_Release1"
     misc_feature      7..442
                       /note="match: GSS: Em:AQ761971"
     repeat_region     245..300
                       /note="14 copies 4 mer tata 73% conserved"
     repeat_region     251..298
                       /note="6 copies 8 mer tatacata 85% conserved"
     repeat_region     256..301
                       /note="23 copies 2 mer at 80% conserved"
     repeat_region     313..989
                       /note="L1ME3A repeat: matches 5406. .6152 of consensus"
     misc_feature      1047..1513
                       /note="match: GSS: Em:AQ819686"
     repeat_region     1908..2053
                       /note="FAM repeat: matches 26. .172 of consensus"
     repeat_region     2276..2574
                       /note="AluSx repeat: matches 9. .303 of consensus"
     repeat_region     2720..2870
                       /note="L2 repeat: matches 2558. .2711 of consensus"
     repeat_region     4330..4437
                       /note="MER94 repeat: matches 10. .121 of consensus"
     repeat_region     4758..4793
                       /note="18 copies 2 mer tt 91% conserved"
     repeat_region     5066..5201
                       /note="MIR repeat: matches 7. .140 of consensus"
     repeat_region     6972..7446
                       /note="L1MEc repeat: matches 278. .780 of consensus"
     repeat_region     7453..8545
                       /note="HERVL repeat: matches 2899. .4004 of consensus"
     repeat_region     8794..9044
                       /note="HERVL repeat: matches 2335. .2589 of consensus"
     repeat_region     9613..9666
                       /note="Alu repeat: matches 1. .53 of consensus"
     repeat_region     9667..9904
                       /note="MLT1D repeat: matches 168. .416 of consensus"
     repeat_region     9916..10048
                       /note="AluJo repeat: matches 1. .136 of consensus"
     repeat_region     10884..11146
                       /note="Tigger3 (Golem) repeat: matches 2771. .3027 of
                       consensus"
     repeat_region     11160..11284
                       /note="Tigger3 (Golem) repeat: matches 1. .125 of consensus"
     repeat_region     11348..11672
                       /note="AluJb repeat: matches 1. .310 of consensus"
     repeat_region     11707..11794
                       /note="MIR repeat: matches 4. .92 of consensus"
     repeat_region     11798..12173
                       /note="L1PA16 repeat: matches 5717. .6155 of consensus"
     repeat_region     12188..12287
                       /note="MIR repeat: matches 138. .235 of consensus"
     repeat_region     12471..16388
                       /note="L1P5 repeat: matches 2293. .6155 of consensus"
     repeat_region     16841..16898
                       /note="L2 repeat: matches 2693. .2750 of consensus"
     repeat_region     18264..18563
                       /note="AluSq repeat: matches 1. .300 of consensus"
     repeat_region     18637..19137
                       /note="L1ME2 repeat: matches 5609. .6131 of consensus"
     repeat_region     19147..19958
                       /note="53 copies 4 mer tata 75% conserved"
     repeat_region     19391..19835
```

```
misc_feature    /note="L1ME2 repeat: matches 5135. .5581 of consensus"
                complement(19871. .20532)
repeat_region   19948. .20130
                /note="match: GSS: Em:AQ283704"
                /note="THE1-INTERNAL repeat: matches 1398. .1580 of
                consensus"
repeat_region   20153. .20452
                /note="THE1C repeat: matches 79. .371 of consensus"
repeat_region   20453. .21157
                /note="L1P repeat: matches 2739. .3657 of consensus"
repeat_region   21330. .21610
                /note="L1M4 repeat: matches 4033. .4299 of consensus"
repeat_region   21611. .21907
                /note="AluY repeat: matches 1. .298 of consensus"
repeat_region   21908. .22023
                /note="L1M4 repeat: matches 3915. .4033 of consensus"
repeat_region   22053. .22549
                /note="LTR44 repeat: matches 1. .519 of consensus"
repeat_region   22814. .22841
                /note="14 copies 2 mer tt 92% conserved"
repeat_region   22922. .22969
misc_feature    24230. .24396
                /note="24 copies 2 mer tt 79% conserved"
repeat_region   24332. .24673
                /note="match: GSS: Em:AQ605495"
                /note="MER65-internal repeat: matches 2142. .2492 of
                consensus"
repeat_region   24674. .24740
                /note="MER51-internal repeat: matches 1581. .1971 of
                consensus"
repeat_region   25057. .25536
                /note="L1PA5 repeat: matches 5655. .6143 of consensus"
repeat_region   25537. .25600
                /note="32 copies 2 mer aa 82% conserved"
repeat_region   25625. .27912
                /note="MER31-internal repeat: matches 3177. .4880 of
                consensus"
repeat_region   28380. .28818
                /note="LTR44 repeat: matches 25. .519 of consensus"
repeat_region   28830. .29139
                /note="AluJb repeat: matches 1. .308 of consensus"
repeat_region   29145. .30686
                /note="L1M4 repeat: matches 2328. .3854 of consensus"
repeat_region   30748. .31111
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   31112. .32639
                /note="THE1B-INTERNAL repeat: matches 1. .1580 of
                consensus"
repeat_region   32640. .33010
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   33591. .33698
                /note="MIR repeat: matches 33. .141 of consensus"
repeat_region   34045. .34274
                /note="MIR repeat: matches 2. .262 of consensus"
repeat_region   35105. .35568
                /note="MLT1C repeat: matches 1. .466 of consensus"
repeat_region   36939. .37047
                /note="Tigger3(Golem) repeat: matches 1. .108 of consensus"
repeat_region   37039. .37745
                /note="Tigger3(Golem) repeat: matches 2277. .3027 of
                consensus"
repeat_region   38735. .39129
repeat_region   41401. 41633
                /note="L2 repeat: matches 1162. .1576 of consensus"
repeat_region   42037. .42383
                /note="MER58A repeat: matches 1. .224 of consensus"
misc_feature    /note="LTR37A repeat: matches 76. .414 of consensus"
                complement(44109. .44432)
misc_feature    /note="match: GSS: Em:B35062"
                complement(44858. .45350)
misc_feature    /note="match: GSS: Em:AQ528681"
                complement(45342. .45500)
                /note="match: GSS: Em:AQ149065"


repeat_region   45561. .46730
                /note="MER11C repeat: matches 1. .1071 of consensus"
repeat_region   47237. .47615
                /note="MSTA repeat: matches 1. .426 of consensus"
misc_feature    47669. .48158
                /note="match: GSS: Em:AQ428522"
repeat_region   47693. .48244
                /note="MER41A repeat: matches 1. .554 of consensus"
repeat_region   48262. .48461
                /note="MIR repeat: matches 53. .262 of consensus"
misc_feature    48927. .49106
                /note="match: GSS: Em:AQ075808"
repeat_region   49149. .49332
                /note="MIR repeat: matches 4. .192 of consensus"
repeat_region   49401. .49551
                /note="L1PA5 repeat: matches 5992. .6143 of consensus"
repeat_region   50633. .50779
                /note="L2 repeat: matches 2541. .2703 of consensus"
repeat_region   51198. .51291
                /note="MIR repeat: matches 14. .115 of consensus"


Query Match             13.6%;  Score 57.2;  DB 9;  Length 204158;
Best Local Similarity   88.6%;  Pred. No. 4e-06;
Matches   62;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        3 GGAGTTTCTTCATTCTGGTAGGTTCGTGGTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 62
            ||||||||||||| |||||||||||| |||||||  ||  ||||||||||||| |||||| |||
Db   155782 GGAGTTTCTTCCTTCTCGGTAGGTTCATGGTCTTCGCGTCTCGCGTCTCAGGAGTGAAGCTGCAGAC 155841

Qy       63 CTCTGCGATG 72
            || |||| ||
Db   155842 CTTTGCGGTG 155851


RESULT 13
AC006456
LOCUS       AC006456               75609 bp    DNA     linear   PRI 26-APR-2003
DEFINITION  Homo sapiens PAC clone RP5-969D4 from 7, complete sequence.
ACCESSION   AC006456
VERSION     AC006456.3  GI:21322189
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 75609)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 75609)
  AUTHORS   Bauer,C., Ozanich,A. and Sandberg,B.
  TITLE     The sequence of Homo sapiens PAC clone RP5-969D4
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 75609)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 75609)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 75609)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 75609)
  AUTHORS   Waterston,R.
```

```
TITLE      Direct Submission
JOURNAL    Submitted (04-JUN-2002) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE  7  (bases 1 to 75609)
AUTHORS    Waterston,R.
TITLE      Direct Submission
JOURNAL    Submitted (26-APR-2003) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT    On Jun 4, 2002 this sequence version replaced gi:4454519.
           ----------------- Genome Center
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu/gsc
           Contact: sapiens@watson.wustl.edu
           ----------------- Summary Statistics
           Center project name: H_DJ0969D04
           -------------------------

NOTICE:    This sequence may not represent the entire insert of this
           clone.  It may be shorter because we only sequence overlapping
           clone sections once, or longer because we provide a small overlap
           between neighboring data submissions.

           This sequence was finished as follows unless otherwise noted:
           all regions were double stranded, sequenced with an alternate
           chemistry, or covered by high quality data (i.e., phred quality >=
           30); an attempt was made to resolve all sequencing problems, such
           as compressions and repeats; all regions were covered by sequence
           from more than one subclone; and the assembly was confirmed by
           restriction digest.

MAPPING INFORMATION:
           The sequence of this clone was established as part of a mapping and
           sequencing collaboration between the NHGRI Chromosome 7 Mapping
           Project (Eric D. Green, Director), John D. McPherson in the
           Department of Genetics (Washington University), and the Washington
           University Genome Sequencing Center. For additional information
           about the map position of this sequence, see
           http://www.nhgri.nih.gov/DIR/GTB/CHR7, send
           mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
           This clone was derived from human PAC library RPCI-4, prepared by
           Pieter de Jong and coworkers at http://www.chori.org using the
           method described by Ioannou et al., Nature Genetics 6:84-9 (1994).
           The library is from one male donor.
           The clone may be obtained either from Genome Systems, Inc.
           (http://www.genomesystems.com) or Research Genetics, Inc.
           (http://www.resgen.com); or from Pieter de Jong.
VECTOR:    pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
           The clone sequenced to the left is H_DJ0558L10, 200 bp overlap.
           Actual start of this clone is at base position 75969 of RP4-558L10.
           actual end is at base position 75609 of RP5-969D4.
FEATURES             Location/Qualifiers
     source          1..75609
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP5-969D4"
                     /clone_lib="RPCI-4"
     repeat_region   601..948
                     /rpt_family="L1"
     variation       921
                     /allele="G"
                     /allele="A"
                     /db_xref="dbSNP:1544590"
     repeat_region   1026..1110
                     /rpt_family="GA-rich"
     variation       1046
```
```
                     /allele="C"
                     /allele="A"
                     /db_xref="dbSNP:1544591"
     repeat_region   1113..1256
                     /rpt_family="A-rich"
     variation       1264
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:2537200"
     repeat_region   1507..1528
                     /rpt_family="AT_rich"
     repeat_region   1928..2057
                     /rpt_family="L1"
     repeat_region   3052..3266
                     /rpt_family="Alu"
     variation       3806
                     /allele="G"
                     /allele="A"
                     /db_xref="dbSNP:2727626"
     variation       3887
                     /allele="T"
                     /allele="C"
                     /db_xref="dbSNP:2727625"
     variation       4632
                     /allele="A"
                     /allele="T"
                     /db_xref="dbSNP:2727624"
     variation       4914
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:957279"
     variation       5067
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:768816"
     variation       5080
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:768817"
     repeat_region   5477..5678
                     /rpt_family="L1"
     repeat_region   5679..5973
                     /rpt_family="Alu"
     repeat_region   5974..6098
                     /rpt_family="L1"
     repeat_region   6101..6431
                     /rpt_family="L1"
     repeat_region   6432..6736
                     /rpt_family="Alu"
     variation       6481
                     /allele="T"
                     /allele="A"
                     /db_xref="dbSNP:726749"
     variation       6598
                     /allele="G"
                     /allele="A"
                     /db_xref="dbSNP:726750"
     repeat_region   6737..6845
                     /rpt_family="L1"
     variation       6749
                     /allele="T"
                     /allele="C"
                     /db_xref="dbSNP:726751"
     variation       7861
                     /allele="C"
                     /allele="A"
                     /db_xref="dbSNP:2727632"
     repeat_region   8094..8163
                     /rpt_family="AcHobo"
     repeat_region   8463..8511
                     /rpt_family="AT_rich"
     variation       10046
                     /allele="A"
```

```
                /allele="G"
                /db_xref="dbSNP:2537199"
repeat_region   10459..10758
                /rpt_family="Alu"
variation       11081
                /allele="G"
                /allele="A"
                /db_xref="dbSNP:887689"
repeat_region   12749..13573
                /rpt_family="ERV1"
repeat_region   13686..13848
                /rpt_family="ERV1"
repeat_region   14311..14339
                /rpt_family="(TTTTG)n"
repeat_region   14516..14702
                /rpt_family="MIR"
repeat_region   16054..16353
```

```
Query Match          13.5%;   Score 57;   DB 9;   Length 75609;
Best Local Similarity  86.3%;   Pred. No. 4.7e-06;
Matches   63;   Conservative   0;   Mismatches  10;   Indels  0;   Gaps   0;

Qy      2  TGGAGTTCTTCATTCGGTAGGTTCGTGGTCCTCTCGGCTTCAGGAATGAAGCTGTAGA 61
           |||||||||||||||||||| |||||| ||||||||||||||||||||||||||||||
Db  12878  TGGAGTTCTTCCTTCGGCGGGGTTCGTGGTCGTCGCTGACTTCAGGAATGAAGCTGCAGA 12937

Qy     62  ACTCTGCGATGCT 74
           ||| ||| ||||
Db  12938  CCTTTGCGGTGTT 12950
```

RESULT 14
AC140889
```
LOCUS       AC140889        172464 bp        DNA       linear   HTG 05-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-171G19, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC140889
VERSION     AC140889.1  GI:28850000
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172464)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 16
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 172464)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (05-MAR-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     ------Genome Center------
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            ------
            Project Information
            Center Project Name: 472192
            Center clone name: RPCI-11_171G19
            ------
            Summary Statistics
            Consensus quality: 167507 bases at least Q40
            Consensus quality: 168805 bases at least Q30
            Consensus quality: 169686 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 171564; sum-of-contigs estimation
            Quality coverage: 5.88 in Q20 bases; agarose-fp estimation
            Quality coverage: 6.0 in Q20 bases; sum-of-contigs estimation
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1522:  contig of 1522 bp in length
            *     1523      1622:  gap of unknown length
            *     1623      3135:  contig of 1513 bp in length
            *     3136      3235:  gap of unknown length
            *     3236      6536:  contig of 3301 bp in length
            *     6537      6636:  gap of unknown length
            *     6637     13115:  contig of 6479 bp in length
            *    13116     13215:  gap of unknown length
            *    13216     23981:  contig of 10766 bp in length
            *    23982     24081:  gap of unknown length
            *    24082     32860:  contig of 8779 bp in length
            *    32861     32960:  gap of unknown length
            *    32961     42311:  contig of 9351 bp in length
            *    42312     42411:  gap of unknown length
            *    42412     68071:  contig of 25660 bp in length
            *    68072     68171:  gap of unknown length
            *    68172    119235:  contig of 51064 bp in length
            *   119236    119335:  gap of unknown length
            *   119336    172464:  contig of 53129 bp in length.
FEATURES             Location/Qualifiers
source               1..172464
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-171G19"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN
```

```
Query Match          13.5%;   Score 57;   DB 2;   Length 172464;
Best Local Similarity  86.3%;   Pred. No. 4.6e-06;
Matches   63;   Conservative   0;   Mismatches  10;   Indels  0;   Gaps   0;

Qy      2  TGGAGTTCTTCATTCGGTAGGTTCGTGGTCCTCTCGGCTTCAGGAATGAAGCTGTAGA 61
           |||||||||||||||||||| |||||| ||||||||||||||||||||||||||||||
Db  21559  TGGAGTTCTTCCTTCGGCGGGGTTCGTGGTCGTCGCTGACTTCAGGAATGAAGCTGCAGA 21618

Qy     62  ACTCTGCGATGCT 74
           ||| ||| ||||
Db  21619  CCTTTGCGGTGTT 21631
```

RESULT 15
AC147084/c
```
LOCUS       AC147084          174895 bp          DNA      linear    HTG 04-NOV-2003
DEFINITION  Pan troglodytes chromosome 7 clone RP43-64A1, *** SEQUENCING IN
            PROGRESS ***, 9 unordered pieces.
ACCESSION   AC147084
VERSION     AC147084.1  GI:38154122
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 174895)
AUTHORS     Wilson,R.K.
TITLE       The sequence of Pan troglodytes clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 174895)
AUTHORS     Wilson,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (04-NOV-2003) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT     --------------- Genome Center ---------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            --------------- Project Information ---------------
```

```
Center project name: C_PT064A01

----------- Summary Statistics -----------
 Sequencing vector: M13; 0%
 Sequencing vector: plasmid; 100%
 Chemistry: Dye-primer ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 171368 bases at least Q40
 Consensus quality: 171979 bases at least Q30
 Consensus quality: 172389 bases at least Q20
-----------------------------------------

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *           1     2971: contig of 2971 bp in length
 *        2972     3071: gap of unknown length
 *        3072    12999: contig of 9928 bp in length
 *       13000    13099: gap of unknown length
 *       13100    30461: contig of 17362 bp in length
 *       30462    30561: gap of unknown length
 *       30562    60139: contig of 29578 bp in length
 *       60140    60239: gap of unknown length
 *       60240    92522: contig of 32283 bp in length
 *       92523    92622: gap of unknown length
 *       92623   170864: contig of 78242 bp in length
 *      170865   170964: gap of unknown length
 *      170965   172086: contig of 1122 bp in length
 *      172087   172186: gap of unknown length
 *      172187   173478: contig of 1292 bp in length
 *      173479   173578: gap of unknown length
 *      173579   174895: contig of 1317 bp in length.

FEATURES             Location/Qualifiers
     source          1..174895
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="7"
                     /clone="RP43-64A1"
     misc_feature    1..2971
                     /note="assembly_name:Contig10"
     misc_feature    3072..12999
                     /note="assembly_name:Contig11"
     misc_feature    13100..30461
                     /note="assembly_name:Contig12"
     misc_feature    30562..60139
                     /note="assembly_name:Contig13"
     misc_feature    60240..92522
                     /note="assembly_name:Contig14"
     misc_feature    92623..170864
                     /note="assembly_name:Contig15"
     misc_feature    170965..172086
                     /note="assembly_name:Contig7"
     misc_feature    172187..173478
                     /note="assembly_name:Contig8"
     misc_feature    173579..174895
                     /note="assembly_name:Contig9"

ORIGIN

Query Match      13.5%;     Score 57;  DB 2;  Length 174895;
Best Local Similarity   86.3%;  Pred. No. 4.6e-06;
Matches  63;  Conservative   0;  Mismatches   10;  Indels    0;  Gaps     0;

Qy        2 TGGAGTTCTTCATTCGGTAGGTTCGTGGTCTTCTCGGCTTCAGGAATGAAGCTTGTAGA 61
            |||||||||||||| ||||||||| ||||||||||| |||||||||||||||| |||||
Db   131899 TGGAGTTCTTCTTCCTTCGTGGTTGGTTCGCGTGGTCTTCGCTGACTTCAGGAATGAAGCTGCAGA 131840

Qy       62 ACTCTGCGATGCT 74
```

```
Db     131839 CCTTCGCGGTGTT 131827
               ||  |||  |||
               CCTTCGCGGTGTT

Search completed: April 25, 2004, 18:58:19
Job time : 1113.58 secs
```

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:     April 25, 2004, 07:00:48 ; Search time 2937.54 Seconds
                                        (without alignments)
                                        16458.288 Million cell updates/sec

Title:      US-10-051-835-15
Perfect score: 1619
Sequence:   1 ttaactacatatcttatgt.........atttattaaaatttaaaa 1619

Scoring table: IDENTITY_NUC
            Gapop 10.0 , Gapext 1.0

Searched:   27513289 seqs, 1493109276 residues

Total number of hits satisfying chosen parameters:    55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
        1: em_estba:*
        2: em_esthum:*
        3: em_estin:*
        4: em_estmu:*
        5: em_estov:*
        6: em_estpl:*
        7: em_estro:*
        8: em_htc:*
        9: gb_est1:*
       10: gb_est2:*
       11: gb_htc:*
       12: gb_est3:*
       13: gb_est4:*
       14: gb_est5:*
       15: em_estfun:*
       16: em_estom:*
       17: em_gss_hum:*
       18: em_gss_inv:*
       19: em_gss_pln:*
       20: em_gss_vrt:*
       21: em_gss_fun:*
       22: em_gss_mam:*
       23: em_gss_mus:*
       24: em_gss_pro:*
       25: em_gss_rod:*
       26: em_gss_phg:*
       27: em_gss_vrl:*
       28: gb_gss1:*
       29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
                 %
Result           Query
No.    Score  Match Length DB ID        ID                  Description
-----  -----  ----- ------ -- --------  --------      --------------------------
  1    615    38.0   772   12 BG431098   BG431098   60249723
c 2    613.8  37.9   745   13 BU681294   BU681294   UI-CF-EC1
c 3    600.2  37.1   685   29 AG165287   AG165287   Pan trogl
c 4    577.6  35.7   696   14 CA312303   CA312303   UI-CF-FN0
```

```
c  5  574.6  35.5  623  10 AW179297  AW179297 PM3-ST011
c  6  569.4  35.2  622  10 AW179289  AW179289 PM3-ST011
   7         34.8  595   9 AI791550  AI791550 ok82d09.y
   8         34.8  560   8 BE866341  BE866341 601678994
   9         32.8  717  10 BG399616  BG399616 602441635
c 10         30.0  735  12 AV719129  AV719129 AV719129
c 11         29.5  486   9 AA205598  AA205598 zq68d01.s
c 12         29.3  618   9 Z98443    Z98443   HS298443 DK
c 13         29.3  580  14 AW969587  AW969587 EST381664
c 14         28.7  475  10 BE564550  BE564550 601343885
c 15         28.5  697  10 AW969584  AW969584 EST381661
c 16         28.4  739  10 CB163886  CB163886 K-EST0224
c 17         28.3  626  14 CA312517  CA312517 UI-CF-FN0
c 18         28.1  498  14 AI862478  AI862478 tdi6h10.x
c 19         26.0  583  14 AW080593  AW080593 xc33g07.x
c 20         25.2  589   9 AW182649  AW182649 xj45a06.x
c 21         24.4  523   9 BG402669  BG402669 602466156
c 22         23.6  524  10 AI934524  AI934524 wp05h10.x
c 23         23.6  903  12 BF208067  BF208067 601862393
c 24         22.9  461   9 BE931720  BE931720 QV4-HT053
c 25         21.8  747  10 BX099236  BX099236 BX099236
c 26         21.7  437  13 BM552062  BM552062 AGENCOURT
c 27         21.6  462  13 AI623825  AI623825 ts23g05.x
c 28         20.1 1041  12 AW390557  AW390557 MR3-ST019
c 29         20.0  451   9 AI628810  AI628810 ty72g04.x
c 30         19.1  590  10 AI922919  AI922919 wn81c03.x
c 31         18.6  326  10 AW662716  AW662716 hi35f03.x
c 32         17.9  430   9 BE145874  BE145874 MR0-HT020
c 33         17.6  412   9 BF755944  BF755944 CM4-CT048
c 34         17.5  406  10 AA625553  AA625553 zu90e03.s
c 35         16.0  399  10 AW511951  AW511951 xu68f08.x
c 36         15.4  322  10 AI500615  AI500615 tn98f03.x
c 37         15.3  387   9 AA862938  AA862938 og98b11.s
c 38         13.5  388  10 AI474738  AI474738 tm36a04.x
c 39         12.6  353   9 AA570161  AA570161 ne57g07.s
c 40         12.2  353   9 AI928006  AI928006 wo92e01.x
c 41         11.8  332  10 AW242799  AW242799 xn26b03.x
c 42         11.0  312   9 AI932894  AI932894 wo39d01.x
c 43         10.4  307  13 BQ427949  BQ427949 AGENCOURT
c 44         10.0  537  14 CA311973  CA311973 UI-CF-FN0
c 45          9.9  269  12 BG433079  BG433079 602496821
```

ALIGNMENTS

```
RESULT 1
BG431098
LOCUS       BG431098                 772 bp    mRNA     linear    EST 14-MAR-2001
DEFINITION  60249723F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4612420 5',
            mRNA sequence.
ACCESSION   BG431098
VERSION     BG431098.1  GI:13337604
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 772)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE     Unpublished (1999)
  JOURNAL   Contact: Robert Strausberg, Ph.D.
  COMMENT   Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1360 row: n column: 05
            High quality sequence stop: 735.
```

```
FEATURES        Location/Qualifiers
     source     1..772
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4612420"
                /lab_host="DH10B (T1 phage-resistant)"
                /clone_lib="NIH MGC 75"
                /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site 1:
                Sfi I (ggccgccctggcc); Site 2: Sfi I (ggccattatggcc); 5' and
                3' adaptors were used in cloning as follows: 5' adaptor
                sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                C, or G and N = A, C, G, or T). Average insert size 1.65
                kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
                by PCR. This library was enriched for full-length clones
                and was constructed by Clontech Laboratories (Palo Alto,
                CA). Note: this is a NIH_MGC Library."
ORIGIN

Query Match      38.0%;   Score 615;   DB 12;   Length 772;
Best Local Similarity   95.4%;   Pred. No. 1.2e-114;
Matches 730;   Conservative   0;   Mismatches   25;   Indels   10;   Gaps   9;
```

```
Best Local Similarity  95.7%;  Pred. No. 2.1e-114;
Matches  717;  Conservative  0;  Mismatches  22;  Indels  10;  Gaps  8;

Qy   871  GACGCTCAGGTTATGATATGTTATGAAAAATAAGCTTCATTATTCTTATAGCTACATC    930
          ||||||||||||||||||||||||||| ||||||||| ||||| |||||||||||||||
Db   745  GACGCTCAGGTTATGATATGTTATGAAAAATAAGCTTCATTATTT-TAGTAGCTACATC    687

Qy   931  CTATTATTCCCTTTAGAAACAAGAATAACATTAGTTTAATAGTTGCCATACTTAGCA    990
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   686  CTATTATTCCCTTTAGAAACAAGAATAACATAGGTTTTAATAGTTGCCATACTTAGCA    627

Qy   991  TTTATCAGTCTAATGAAACCAATATTGAATCTCGATAAATATTTCTGATGTTACTAG   1050
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   626  TTTATCAGTCTAATGAAACCAATATTGAATCTCGATAAATATTTCTGATGTTACTAG    567

Qy  1051  CTATGGGAAATTAGAACTGCCACAACCCTGACATTACTAAGTGGAAATGTTAGGATTTT   1110
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   566  CTATGGGAAATTAGAACTGGCACAACCCTGACATTACTAAGTGGAAATGTTAGGATTTT    507

Qy  1111  CGGCATCGCATGTTAGAATCTCTAAAATTTAAACATTCCTGTTAAATGACTAAGGTTTGC   1170
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   506  CGGCATCGCATGTTAGAATCTCTAAAATTTAAACATTCCTGTTAAATGACTAAGGTTTGC    447

Qy  1171  TTTTATCAATATGAATTCTGAAGGCCAATATCATTACCATTAACCATATGAAAGCTTTAATT   1230
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   446  TTTTATCAATATGAATTCTGAAGGCCAATATCATTACCATTAACCATATGAAAGCTTTAATT    387

Qy  1231  CCTAAAAAATAGTTTTAGAGATATTCAAGCAATGCTCTCCTAATATCCATACGCAAGTGTG   1290
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   386  CCTAAAAAATAGTTTTAGAGATATTCAAGCAATGCTCTCCTAATATCCATACGCAAGTGTG    327

Qy  1291  TTTATGACACAAAATTCACTAGTCTGGTTTAAAAATGAAATCTTTATATTGACTGACTGGTGTC   1350
          ||||||||||||||||||||||||| ||| |||||||||| |||||||||||| ||||||
Db   326  TTTATGACACAAAATTCACTAGTCT-GTTAAAAATGAATTCTTTATATTGACT-GGTGTT    269

Qy  1351  CCACATATTCCAGTAATTTCTGTTATGAGAGGACTTGAAATAGCAAATTGCCCCACAGT   1410
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   268  CCACATATTCCAGTAATTTCTGTTATGAGAGGACTTGAAATAGCAAATTG-CCACACAGT    210

Qy  1411  TAACTGGATAGATCACGTACGTGGTGATCATAACCACTTGGTACTACACCCAGAAACTCA   1470
          ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   209  TAACTGGATAGACCACGTACGTGGTGATCATAACCACTTGGTACTACACCCAGAAACTCA    150

Qy  1471  AAATTGTCTTTCTCCTGATGAGATATGGGTGTCCTTTGTACGTCTTAGGGCCTAGGGT   1530
          ||||||| ||||||||||||||||| |||||||||||||||||||||||||   ||||
Db   149  AAATTGTC-TTTCTCCTGATGAGATAT-GGGTGTCCTTTGTACGTCTTAGGGCC---TAGG     95

Qy  1531  ACCCAAGTGAAGTGAATATATAAGACAAAATGTGTTTGTATCCGAGTCTTCCTGTCATTG   1590
          | ||| ||| ||||| |  | ||| ||||||||||||||||||||||||||||||||||
Db    94  TAACCAGTGGAGTGATTGATTATTAGC-AAATGTGTTTGTATCCGAGTCTTCCTGTCATTG     36

Qy  1591  TAATAAAAAATTTATTTAATTTAAAAATTAAAA   1619
          |||||||||||||||||||||||||||  ||||
Db    35  TAATAAAAAATTTATTTAATTTAAAAAAAAAAAAA      7
```

```
RESULT 3
AG165287/c
LOCUS       AG165287              685 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-032L08.T7, genomic survey
            sequence.
ACCESSION   AG165287
VERSION     AG165287.1  GI:1669965
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       BAC end sequences of Library RPCI-43
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 685)
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
            PRIMERS
            Sequencing: T7
            LIBRARY
            Vector   : pBACe3.6
            R.Site 1 : EcoRI
            R.Site 2 : EcoRI
FEATURES             Location/Qualifiers
     source          1..685
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-032L08.T7"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN
```

```
Query Match         37.1%;  Score 600.2;  DB 29;  Length 685;
Best Local Similarity  96.1%;  Pred. No. 1.2e-111;
Matches  647;  Conservative  0;  Mismatches  23;  Indels  3;  Gaps  3;

Qy   660  CCTCCTTTAGAAACAGTCAAAAGCCACTCAAACAACAGTTCAAACAAGTTCAAAATAAGGAAGGTAGCA    719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   684  CCTCCTTTAGAAACAGTCAAAAGCCACTCAAACAACAGTTCAAACAAGTTCAAAATAAGGAAGGTAGCA    625

Qy   720  AGTTAGGCGATGGATATATATTTTCTTGGCTTGTGTGTATA-CCAGTTGTCAAGGACATTAT    779
          |||||||||||||||||||||||| ||| |||||||||||| |||||||||||||||||||
Db   624  AGTTAGGCGATGGATATATATTTTTCTT-GCTTGTGTGTATA-CCAGTTGTCAAGGACATTAT    567

Qy   780  AAGGACTTCCAAAAGCATTTTGAAGAATGGCAATATCAAAATAAGTGAAATAATAAGTGTATGTCCTCCAA    839
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||| |||||||
Db   566  AAGGACTTCCAAAAGCATTTTGAAGGATGGCAATATCAAAATAAGTGAAATAATAAGTGTATGTCTCTCAA    507

Qy   840  TGAGGCATTTTAATTGTTACAATCTATTGGACGCTCAGTTGATGATATGTTTATGAAA    899
          ||||||||||||||||||  ||  |||| ||||| ||| || |||||||||||||||||
Db   506  TGAGGCATTTTAATTGTTAAAAACCATTTGGACCCTCCAGTTATGATATGTTTATGAAA    447

Qy   900  AATAAGCTTCATTATTTCTTATAGCTACATCCTATTATTCCTTTAGAAACAAGAATAA    959
          |||||||||||||||| |||| |||||||||||||||||||||||||||||||||||
Db   446  AATAAGCTTCATTATTT-TTATCGCTACATCCTATTATTCCTTTAGAAACAAGAATAA    388

Qy   960  CAATAAGTTTTATAGTTGCCATACTTAGCATCATTAGTGGTCTATGGGAAATGAAACCAATATTGA   1019
          |||||||||||||||||||||||||||||||||||||   | |||||||||||||||||||||
Db   387  CAATAAGTTTTATAGTTGCCATACTTAGCATCATTT-ATCAGTCTAATGAAACCAATATTGA    328

Qy  1020  ATCTTCGATAAATATTTCTGATGTTACTAGGTCTATGGGAAATTAGAACTGGCACACACCT   1079
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   327  ATCTTCGATAAATATTTCTGATGTTACTAGGTCTATGGGAAATTAGAACTGGCAAAACCCT    268

Qy  1080  GACATTACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGGAATCTCTAAAATT   1139
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   267  GACATTACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGGAATCTCTAAAATT    208

Qy  1140  TAAACATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATATGAATTCTGAAGGCCAAT   1199
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   207  TAAACATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATATGAATTCTGAAGGCCAAT    148

Qy  1200  ATCATACCATTAACTAACTATGAAAGCTTTTAATTCCTAAAAAATAGTTTTAGAGATATTCAAGC   1259
          ||||||| |||||| ||||| |||||||||||||||||||||||||||||||||||||||||||
Db   147  ATCATACCCTAACTTAACTGAAAGCTTTTAATTCCTAAAAAATAGTTTTAGAGATATTCAAGC     88
```

```
Qy  1260 AATGCTCCTAATATCCATACGCAAGCTGTGTTTATGACACAAATTCACTAGTCCTGGTTT 1319
         |||||||||||||||||||||||||||||||||||||||||||||           ||
Db    87 AATGCTCCTAATATCCATACGCAAGCTGTGTTTATGACACGAATTCACTAGTCGTGTTTA 28

Qy  1320 AAAAATGAAATCT 1332
         ||||      | ||
Db    27 AAAATGGGATTCT 15
```

RESULT 4
CA312303/c
LOCUS   CA312303   696 bp   mRNA   linear   EST 04-NOV-2002
DEFINITION UI-CF-FN0-afm-n-21-0-UI.s1 UI-CF-FN0 Homo sapiens cDNA clone
            UI-CF-FN0-afm-n-21-0-UI 3', mRNA sequence.
ACCESSION   CA312303
VERSION   CA312303.1 GI:24530401
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 696)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE   Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   MEDLINE   97044477
   PUBMED   8889548
  COMMENT   Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com) or from Open Biosystems
            (www.openbiosystems.com).
            The following repetitive elements were found in this cDNA
            sequence: 1-35, >AT rich#Low_complexity
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES   Location/Qualifiers
     source   1..696
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-FN0-afm-n-21-0-UI"
                     /tissue_type="Human Lung Epithelial cells"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-CF-FN0"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-FN0 is a subtracted cDNA library derived from two
                     normalized Human lung epithelial cell libraries (EN1 and
                     DU1) The library was subtracted according to according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. For additional information, contact:
                     bento-soares@uiowa.edu
                     TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
                     TAG_LIB=UI-CF-FN0
                     TAG_SEQ=GGCTGTAGGC"
ORIGIN

Query Match   35.7%; Score 577.6; DB 14; Length 696;
Best Local Similarity   95.6%; Pred. No. 4.7e-107;
Matches 668; Conservative 0; Mismatches 22; Indels 9; Gaps 7;

```
Qy   921 TAGCTACACCTATTATTCCCTTTAGAACAGAATAACAATAAGTTTAATAGTTGCC 980
```

```
Db   696 TAGCTACATCCTATTATTCCCTTTNAGAACAGAATAACAAGAATAAGTTTAATAGTAGTTGCC 637

Qy   981 AATACTTAGCATTTATCAGGTCTAATGAAACCAATATTGAATCTCTGATAAATATTTTCTG 1040
         |||||||||||| |||||||||||||||||||||||||| |||||||||||||||||||
Db   636 AATACTTAGCATGTATCAGGTCTAATGAAACCAATATATTGAATCTCTGATAAATATTTTCTG 577

Qy  1041 ATGTTACTAGCTATGGGAAATAGAACTGGCACAACCCTGACATTACTAAGTGGAAATGT 1100
         ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   576 ATGTTACTAGCTATGNGAAATAGAACTGGCACAACCCTGACATTACTAAGTGGAAATGT 517

Qy  1101 TAGGATTTTCGGCATCCGCATGTTAGAATCTCTAAAATTAAACATTCCTGTTAAATGAC 1160
         |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db   516 TAGGATTTTCGGCATCCGCATGTTAGAATCTCTAAAATNTAAACATTCCTGTTAAATGAC 457

Qy  1161 TAAGGTTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCATTACCATGTATGAAA 1220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   456 TAAGGTTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCATACATACCATACTAATGAAA 397

Qy  1221 GCTTTAATTCCTAAAAAATAGTTTTAGAGATATTCAAGCAAGTGCTCTCCTAATATCCATA 1280
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   396 GCTTTAATTCCTAAAAAATAGTTTTAGAGATATTCAAGCAAGTGCTCTCCTAATATCCATA 337

Qy  1281 CGCAAGTGTGTTTATGACACAAATTCACTAGTCTGGTTTAAAAATGAAATCTTTATATTG 1340
         |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   336 CGCAAGTGTGTTTATGACACAAATTCACTAGTCT-GTTAAAAATGAATTCTTATATTG 278

Qy  1341 ACTGGGTGCCCCACATATTCCAGTAATTTCTGTTATGAGAGGACTTGAAATAGCAAATTG 1400
         ||||||||||       ||||||||||||||||||||||||||||||||||||||||||||
Db   277 ACT-GGTGTTCCACATATTTCAGTAATTTCGTTATGAGAGGACTTGAAATAGCAAATTG 219

Qy  1401 CCCACACAGTTAACTGGATAGATCACGTACGTGGTGATCATAACCACTGGTACTACACC 1460
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   218 -CCACAGTTAACTGGATAGACCACGTACGTGTGTGATCATAACCACTGGTACTACACC 160

Qy  1461 CAGAAACTCAAAATTGTCTTTCTCCTGATGAGATACGGGGTGTCCTTTTGTACGTCTCAG 1520
         |||||||||||||||| ||||||||||||||||||||| |||||||||||||||||||||
Db   159 CAGAAACTCAAAATTGTC-TTTCTCCTGATGAGATAT-GGGTGTCCTTTTGTACGTCTCAG 102

Qy  1521 GGCCTAGGGTACCCAAGTGAAGTGAGATATATAAGCAAAATGTGTTTGTATCCAGAGTCTT 1580
         |||||||          |||||||||||||||||||||||  ||||||||||||||||||||
Db   101 GCC---TAGGTAACCAGTGGAGTGGATTGATTATTAGC-AAATGTGTTTGTATCCAGAGTCTT 46

Qy  1581 CCTGTTCATTGTAATAAAAAATTATTTAAAAATTTAAA 1619
         ||||||||||||||||||||||||||||||||||||||
Db    45 CCTGTCATTGTAATAAAAAATTTATTTAAAAAAAAAAAA 7
```

RESULT 5
AW179297/c
LOCUS   AW179297   623 bp   mRNA   linear   EST 16-NOV-1999
DEFINITION PM3-ST0115-100999-001-C07 ST0115 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW179297
VERSION   AW179297.1 GI:6445334
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 623)
  AUTHORS   Simpson A.J.G.
  TITLE   HCGP http://www.ludwig.org.br/ORESTES.
            The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL

ACCESSION   AW179289
VERSION   AW179289.1   GI:6445326
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 622)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE   The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM3&t2=PM3-ST0115-
             100999-001-B01&t3=1999-09-10&t4=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 615.
FEATURES   Location/Qualifiers
     source   1..622
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="ST0115"
                     /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

Query Match   35.2%;   Score 569.4; DB 10; Length 622;
Best Local Similarity 97.4%;   Pred. No. 2.2e-105;
Matches 609;   Conservative   0; Mismatches   13;   Indels   3;   Gaps   3;

Qy   241 ATAAAGCATTAAATATATGATATATAGCTATATCTATGTATGTATCTAACACAGAGAAGTTC 300
          |||||||||||| |||||||||||||||||||||||||||||||||||||| |||||||||
Db   622 ATAAAGCATTAAATATANATATATGATATATAGCTATATCTATGTATGTATCTAACAGAAGTTC 563

Qy   301 AAGTCACTTCAATTAAAGAAACATTTTTGAGCATGGGACCAGCCTCAGGTCTTATGCTGG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   562 AAGTCACTTCAATTAAAGAAACATTTTTGAGCATGGGACCACCCTCAGGTCTTATGCTGG 503

Qy   361 GATGCAGTAGACAGGAGATGGGGAAATTAGAAAAGAGAACTGTGTAATTGAAATGACGTG 420
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   502 GATGCAGTAGACAGGAGATGGGGAAATTAGAANAGAGAAACTGTGTAATTGAAATGACGTG 443

Qy   421 GGCTGCACCCTTAAGGAACTTATAATTAATGATGATCTGAATAAACATACCAGGATAAAG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   442 GGCTGCACCCTTAAGGAACTTATAATTAATGATGATCTGAATAAACATACCAGGATAAAG 383

Qy   481 ATGTCAAATGAGTGTGACTCCCCTAAAGTAGAATTAAAGTGTGCATTCTTGTTTCCTAAA 540
          |||||||||||||||||||||| ||||||| |||||||||||||||| ||||||||||||
Db   382 ATGTCAAATGAGTGTGACTCCCCTAAAGTAGAATTAAAGTGTGCATTCTTGTTTCCTAAA 323

Qy   541 ATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAACTAACATTAGCATTA 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   322 ATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAACTAACATTAGCATTA 263

Qy   601 ATATGAATAATGCATGGAAAATATTATCCTTTGAAAACTGATTGATAAATATATTCCCC 660
          ||||||||||| |||||||||||||||||||||||||||||||||| ||||||||||||
Db   262 ATATGAATAAT-CATGGAAAATATTATCCTTTGAAAACTGATTGATAAATATATTCCCC 204

             (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM3&t2=PM3-ST0115-
             100999-001-C07&t3=1999-09-10&t4=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 609.
FEATURES   Location/Qualifiers
     source   1..623
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="ST0115"
                     /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

Query Match   35.5%;   Score 574.6; DB 10; Length 623;
Best Local Similarity 97.9%;   Pred. No. 1.9e-106;
Matches 613;   Conservative   0; Mismatches   10;   Indels   3;   Gaps   3;

Qy   240 TATAAAGCATTAAAATATATGATATATAGCTATATCTATGTATGTATCTAACACAGAGAAGTT 299
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   623 TATAAAGCATTAAAATATATGATATATAGCTATATCTATGTATGTATCTANCACAGAGAAGTT 564

Qy   300 CAAGTCACTTCAATTAAAGAAACATTTTTGAGCATGGGACCAGCCTCAGGTCTTATGCTG 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   563 CAAGTCACTTCAATTAAAGAAACATTTTTGAGCATGGGACCACCCTCAGGTCTTATGCTG 504

Qy   360 GGATGCAGTAGACAGGAGATGGGGAAATTAGAAAAGAGAACTGTGTAATTGAAATGACGT 419
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   503 GGATGCAGTAGACAGGAGATGGGGAAATTAGAANAGAGAACTGTGTAATTGAAATGACGT 444

Qy   420 GGGCTGCACCCTTAAGGAACTTATAATTAATGATGATCTGAATAAACATACCAGGATAAA 479
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   443 GGGCTGCACCCTTAAGGAACTTATAATTAATGATGATCTGAATAAACATACCAGGATAAA 384

Qy   480 GATGTCAAATGAGTGTGACTCCCCTAAAGTAGAATTAAAGTGTGCATTCTTGTTTCCTAA 539
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   383 GATGTCAAATGAGTGTGACTCCCCTAAAGTAGAATTAAAGTGTGCATTCTTGTTTCCTAA 324

Qy   540 AATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAGTTAGAAACTAACATAGCATT 599
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   323 AATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAGTTAGAAACTAACATAGCATT 264

Qy   600 AATATGAATAATGCATGGAAAATTATTATCCTTTGAAAACTGATTGATAAATATATTCCC 659
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   263 AATATGAATAAT-CATGGAAAATTATTATCCTTTGAAAACTGATTGATAAATATATTCCC 205

Qy   660 CCTCCTTTAGAAACAGTCAAAAGCCACTTCAAACAAGTTTCAAATAAAGGAAGGTAGCA 719
          |||||||||||||||||||||| ||||| |||||||||||||||||||||||||||||
Db   204 CCTCCTTTAGAAACAGTCAAAAGCCACTTCAAACAAGTTTCAAATAAAGGAAGGTAGCA 145

Qy   720 AGTTAGGCGGATGGATTATATTTTCTTGGCGTTGTTGTATACCCATTGGCCAGGGCCTTTAT 779
          |||||||||||||||||||||||||| |||||| ||| | |||||||||||||||||||
Db   144 AGTTAGGCGGATGGATTATATTTTCTT-GCTTGTTGTATA-CCAGTTGTCAAGGACATTAT 87

Qy   780 AAGGACTCCCAAAAAGCATTTTGAAGAATGGCAATATCAAATAAGTGTANGTCCTCTCAAA 839
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   86 AAGGACTCCCAAAAAGCATTTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCTCAAA 27

Qy   840 TGAGGCATTTTAATTGTTACAATCT 865
          ||||||||| |||||| |||||||
Db   26 TGAGGCATTTTTAATTGTTAAAATCT 1

RESULT 6
AW179289/c
LOCUS   AW179289   622 bp   mRNA   linear   EST 16-NOV-1999
DEFINITION PM3-ST0115-100999-001-B01 ST0115 Homo sapiens cDNA, mRNA sequence.

```
Qy  661 CTCCCTTAGAAACAGTCAAAAGCCACTTCAAACAAGTTCAAAAATAAAGGAAGGTAGCAA 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  203 CTCCCTTAGAAACAGTCAAAAGCCACTTCAAACAAGTTCAAAAATAAAGGAAGGTAGCAA 144

Qy  721 GTTAGGGCGATGGATTATATTTCTTGGCTTGTTGTATACCCATTGGCCAGGGCCTTTATA 780
        |||||||||||||||||||||||||  |||||||||||  |  ||| || |||| |||||
Db  143 GTTAGGGCGATGGATTATATTTCTT-GCTTGTTGTATA-CCAGTTGTCAAGAACATTATA 86

Qy  781 AGGACTCCCAAAAGCATTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTTCTCAAAT 840
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 AGGACTCCCAAAAGCATTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTTCTCAAAT 26

Qy  841 GAGGCATTTTAATTGTTACAATCT 865
        ||||||||||||||||||   ||||
Db   25 GAGGCATTTTAATTGTTAAAATCT 1
```

RESULT 7
AI791550

LOCUS   AI791550   595 bp   mRNA   linear   EST 13-DEC-1999
DEFINITION ok82d09.y5 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1520465 5',
            mRNA sequence.
ACCESSION   AI791550
VERSION   AI791550.1 GI:5339266
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 595)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Other_ESTs: ok82d09.x5
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            This read is a RESEQUENCE of a previously sequenced human clone
            Original clone citation: see original entry for original citation
            information
            This 5' resequenced clone has no previous 5' data to verify this
            new read against.
            Insert Length: 3866   Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 459.
FEATURES   Location/Qualifiers
     source   1..595
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1520465"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Kid3"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo. "
ORIGIN

```
Query Match             34.8%;  Score 564;  DB 9;  Length 595;
Best Local Similarity   95.8%;  Pred. No. 2.8E-104;
Matches  564;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy   23 ATAAATATTGCCTAATCCATAGAAAAAAGGATATAAAGTATTAAATATGTGATATAGC 82
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATAAATATTGCCTAATCCATAGAAAAAAGGATATAAAGTATTAAATATGTGATATAGC 60

Qy   83 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATAGGGAAGTTCAAGTCACTTCAATTGA 142
                                      ||||||||||||||||||||||||||||||
Db   61 TATATCTATCTATGTATCTAATAGGGAAGTTCAAGTCACTTCAATTGAGAAAACAT 120

Qy  143 ATCTCTGAGCATAGAGAGCAGCCTCAGGTCCTATGTGGGATGCAGTGGACAGGAGAGGG 202
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ATCTCTGAGCATAGAGAGCAGCCTCAGGTCCTATGTGGGATGCAGTGGACAGGAGAGGG 180

Qy  203 GAAATTAGAAAAGAGAACTATATAATTGAAAAACGGATATAAAGCATTAAATATATGATA 262
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GAAATTAGAAAAGAGAACTATATAATTGAAAAACGGATATAAAGCATTAAATATATGATA 240

Qy  263 TATAGCTATATCTATGTATGTATCTAACAGAGAAGTTCACTCACTTCAATTAAAGAAAC 322
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TATAGCTATATCTATGTATGTATCTAACAGAGAAGTTCACTCACTTCAATTAAAGAAAC 300

Qy  323 ATTTTTGAGCATGGGACCAGCCTCAGGTCTTATGCTGGGATGCAGTAGACAGGAGATGGG 382
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 ATTTTTGAGCATGGGACCAGCCTCAGGTCTTATGCTGGGATGCAGTAGACAGGAGATGGG 360

Qy  383 GAAATTAGAAAAGAGAACTGTGTAATTGAAATGACGTGGGCTGCACCCTTAAGGAACTTA 442
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GAAATTAGAAAAGAGAACTGTGTAATTGAAATGACGTGGGCTGCACCCTTAAGGAACTTA 420

Qy  443 TAATTAATGATGATCTGAATAAACATACCAGGATAAAGATGTCAAATGAGTGTGACTCCC 502
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 TAATTAATGATGATCTGAATAAACATACCAGGATAAAGATGTCAAATGAGTGTGACTCCC 480

Qy  503 TTAAAGTAGATTAAAGTGTGCATTCTTTGTTTCCTAAAATATGATTTTTACTGCTTGAAAT 562
         |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  481 TTAAAGTAGATTAAAGTGTGCATTCTTTGTTTCCTAAAATATGATTNTNTACTGCTTGAAAT 540

Qy  563 TACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTAATATGAATAAT 611
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTAATATGAATAAT 589
```

RESULT 8
BB866341

LOCUS   BB866341   717 bp   mRNA   linear   EST 27-SEP-2000
DEFINITION 601678994F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3961510 5',
            mRNA sequence.
ACCESSION   BB866341
VERSION   BB866341.1 GI:10315221
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 717)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM845 Plate: 1 column: 23
            High quality sequence stop: 630.
FEATURES   Location/Qualifiers
     source   1..717

Qy   1495 TATGGG 1500
           ||||||
Db   712 TATGGG 717

RESULT 9
BG399616
LOCUS   BG399616   735 bp   mRNA   linear   EST 12-MAR-2001
DEFINITION 6024416351 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4557321 5',
            mRNA sequence.
ACCESSION   BG399616
VERSION   BG399616.1 GI:13292983
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 735)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1261 row: f column: 10
            High quality sequence stop: 575.
FEATURES   Location/Qualifiers
     source   1..735
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4557321"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_75"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3', and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match   32.8%;   Score 531.8;   DB 12;   Length 735;
Best Local Similarity 89.9%;   Pred. No. 9e-98;
Matches 612;   Conservative 0;   Mismatches 61;   Indels 8;   Gaps 5;

Qy   1 TTAATCACATATCTTATGTATATAAATATTGCCTAATCCATAGAAAAAAAGATATAAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   37 TTAATCACATATCTTATGTATATAAATATTGCCTAATCCATAGAAAAAAAGATATAAAG 96

Qy   61 TATTAAAATATGTGATATATAGCNNNNNNNNNNNNNNNNNNNNNATAGGGAAGTTCAA 120
           ||||||||||||||||||||||||
Db   97 TATTAAAATATGTGATATATAGCTATATCTATCTATCTATCTAATAGGGAAGTTCAA 156

Qy   121 GTCACTTCAATTGAAGAAACATATCTTCGAGCATAGGAGCAGCCTCAGGTCCTATGGTGG 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   157 GTCACTTCAATTGAAGAAACATATCTTCGAGCATAGGAGCAGCCTCAGGTCCTATGGTGG 216

Qy   181 GATGCAGTGGACAGGAGGGGGAAATTAGAAAGAAGAACTATATATTGAAAAAGGGAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   217 GATGCAGTGGACAGGAGGGGGAAATTAGAAAGAAGAACTATATATTGAAAAAGGGAT 276

Qy   241 ATAAAGCATTAAATATGATATAGCTATATCATCTATGTATGTATCTCAACAGAGAAGTTC 300

                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3961510"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_53"
                     /note="Organ: bladder; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI
                     (ggccattatggcc); Double-stranded cDNA was prepared from
                     cell line RNA. 5' and 3' adaptors were used in cloning as
                     follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3', and
                     3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.55
                     kb (range 0.9-4.0 kb). 15/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA)."

ORIGIN

Query Match   34.6%;   Score 560;   DB 10;   Length 717;
Best Local Similarity 93.0%;   Pred. No. 1.7e-103;
Matches 675;   Conservative 0;   Mismatches 40;   Indels 11;   Gaps 8;

Qy   777 TATAAGGACTCCCAAAAGCATTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCTC 836
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 TATAAGGACTCCCAAAAGCATTTCGAAGGATGGCAATATCAAATAAGTGTATGTCCTCTC 60

Qy   837 AAATGAGGCATTTTAATTGTTACAATCTATTTGGACGCTCAGTTGATGATATGTTTATG 896
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAATGAGGCATTTTAATTGTTAAAATCTATTTGGACGCTCAGTTATGATATGTTTATG 120

Qy   897 AAAAATAAGCTTCATTATTTCTTATAGCTACATCCTATTATTCCCTTTTAGAAACAAGAA 956
           ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   121 AAAAATAAGCTTCATTATTTCTTATATT-TTATAGCTACATCCTATTATTCCCTTTTAGAAACAAGAA 179

Qy   957 TAACAATAAGTTTTAATAGTTGCCATACTTAGCATTTATCAGTTCTAATGAAACCAATAT 1016
           |||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   180 TAACAATAAGTTT---TAGTTGCCATACTTAGCATTTATCAGTTCTAATGAAACCAATA- 235

Qy   1017 TGAATCTCTGATAAATATTTCTGATGTTACTAGCTATGGTATGGGAAATTAGAACTGGCACAAC 1076
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   236 TGAATCTCTGATAAATATTTCTGATGTTACTAGCTATGGGAAATTAGAACTCGGCACAAC 295

Qy   1077 CCTGACATTACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAAATCTCTAAA 1136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   296 CCTGACATTACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAAATCTCTAAA 355

Qy   1137 ATTTAAAACATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATAATGAATTCTGAAGGCC 1196
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   356 ATTTAAACATTCCTGTTAAATGACTAAGG-TTGCTTTTATCAATAATGAATTCTGAAGGCC 414

Qy   1197 AATATCATACCATTAACTATGAAAGCTTTTAATTCCTAAAAAATAGTTTTAGAGATATTCA 1256
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   415 AATATCATACCATTAACTATGAAAGCTTTTAATTCCTAAAAAATAGTTTTAGAGATATTCA 474

Qy   1257 AGCAATGCTCTCCTAATATCCATACGCAA--GTGTGTTATGACACAAATTCACTAGTCT 1314
           ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   475 AGCAATGCTCTCCAATATCCATACGCAAAGTGTGGTTTATGACACAAATTCACTAGTCT 534

Qy   1315 GGTTTAAAAATGAAATCTTTATATTGACTGGGGTGTCCCACATATTCCAGTAATTCTGTT 1374
           ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   535 -GTTTAAAAATGAAATTCTTTATA-TGACTGGTTGTGTTCCCACATATTTCAGTAATTTCTGTT 592

Qy   1375 ATGAGAGGACTTGAAATAGCAAATTGCCCACAGTTAACTGGATAGATCACGTACGTGG 1434
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   593 ATGAGAGGACTTGAACATAGCAAATTGCCACACAGTTAACCTGGACCACGTACGTGG 652

Qy   1435 TGATCATAACCACTTGGTACTACACCCAGAAACTCAAAATTGTCTTTCTCCCTGATGAGA 1494
           ||||||| ||||| |||||||||||||||||||||||||||||||||||||||||
Db   653 TGATCATACCCAC-TGGTAATACACCCAGAAAATCAAATGGCCTTTCTCCTGGTGAGA 711

Best Local Similarity 96.0%; Pred. No. 1.8e-88;
Matches 574; Conservative 0; Mismatches 15; Indels 9; Gaps 7;

```
Qy     997 AGGTCTAATGAAACCAATATTGAATCTCTGATAAATATTTTCTGATAGTTACTAGTAGCTATGG 1056
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     589 AGGTCTAATGAAACCAATATTGAATCTCTGATAAATATTTTCTGATAGTTACTAGCTATGG 530

Qy    1057 GAAATTAGAACTGGCACACCCTGACATTACTAAGTGAAATGTTAGGATTTTTCGGCAT 1116
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     529 GAAATTAGAACTGGCACACCCTGACATTACTAAGTGAAATGTTAGGATTTTTCGGCAT 470

Qy    1117 CGCATGTTAGAATCTCTAAATTTAAACATTCCTGTTAAATGACTAAGGTTGCTTTTAT 1176
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     469 CGCATGTTAGAATCTCTAAATTTAAACATTCCTGTTAAATGACTAAGGTTGCTTTTAT 410

Qy    1177 CAATATGAATTCTGAAGGCCAATATCATACCATTAACTATGAAAGCTTTTAATTCCTAAA 1236
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     409 CAATATGAATTCTGAAGGCCAATATCATACCATTAACTATGAAAGCTTTTAATTCCTAAA 350

Qy    1237 AATAGTTTTAGAGATATTCAAGCAATGCTCTCCTAATATCCATACGCAAGTGTGTTTATG 1296
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     349 AATAGTTTTAGAGATATTCAAGCAATGCTCTCCTAATATCCATACGCAAGTGTGTTTATG 290

Qy    1297 ACACAAATTCACTAGTCTGGTTTAAAAATGAAATCTTTATATTGACTGGGTGTCCCACAT 1356
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     289 ACACAAATTCACTAGTCT-GTTTAAAATGAATTCTTTATATTGACT-GGTGTTCCACAT 232

Qy    1357 ATTCCAGTAATTCTGTTATGAGAGGACTTGAAATAGCAAATTGCCCACACAGTTAACTG 1416
           ||||  |||||||||||||||||||||||||||||||||||||||  |||||||||||||
Db     231 ATTTCAGTAATTTCTGTTATGAGAGGACTTGAAATAGCAAATTG-CCACACAGTTAACTG 173

Qy    1417 GATAGATCACGTACGTGGTGATCATAACCACTTGGTACTAACCACTTGGTACTACACCCAGAAACTCAAAATTG 1476
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172 GATAGACCACGTACGTGGTGATCATAACCACTTGGTACTACACCCAGAAACTCAAAATTG 113

Qy    1477 TCTTTTCTCCTGATGAGATATGGGGGTGTCCTTTTGTACGTCTCAGGGGCTCAGGGTACCCAA 1536
           || |||||||||||||||||||||||| ||||||||||||||||||||||   -TAGGTAACCA
Db     112 TC-TTTCTCCTGATGAGATAT-GGGTGTCCTTTTGTACGTCTCAGGCC---TAGGTAACCA 58

Qy    1537 GTGAAGTGAATATATAAGCAAAATGTGTTTGTATCCAGAGTCTTCCTGTCATTGTAAT 1594
           |||||||||||||||||||||||   ||||||||||||||||||||||||||||||| 1
Db      57 GTGGAGTGATTATATTAGC-AAATGTGTTTGTATCCAGAGTCTTCCTGTCATTGTAAT 1
```

RESULT 11
AA205598/c
LOCUS   AA205598   618 bp   mRNA   linear   EST 27-JAN-1997
DEFINITION   zg68d01.s1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA
              clone IMAGE:646753 3', mRNA sequence.
ACCESSION   AA205598
VERSION   AA205598.1   GI:1803606
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 618)
AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
              Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
              Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
              Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
              Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
              Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
              and Marra,M.
TITLE   Generation and analysis of 280,000 human expressed sequence tags
JOURNAL   Genome Res. 6 (9), 807-828 (1996)
MEDLINE   97044478
PUBMED   8889549
COMMENT   Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810

```
Db     277 ATAAAGCATTAAAATATATGATAATAGCTATATCTATGTATGTATGTATCTAACACAGAGAAGTTC 336
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     301 AAGTCACTTCAATTAAAGAAAACATTTTGAGCATGGGACCCAGCCTTCAGTCTTTATGCTGG 360
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     337 AAGTCACTTCAATTAAAGAAAACATTTTGAGCATGGGACCCAGCCTTCAGTCTTTATGCTGG 396

Qy     361 GATGCAGTAGACAGAGAGGATCGGGAAATTAGAAAAGAGAACTGTGTAATTGAAATGACGTG 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     397 GATGCAGTAGACAGAGAGGATCGGGAAATTAGAAAAGAGAACTGTGTAATTGAAATGACGTG 456

Qy     421 GGCTGCACCCTAAGGAACTTATATATTAATAATGATGATCTGAATAAACATACCAGGATAAAG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     457 GGCTGCACCCTAAGGAACTTATATATTAATAATGATGATCTGAATAAACATACCAGGATAAAG 516

Qy     481 ATGTCAAATGAGTGTGACTCCCTAAAGTAGAATTAAAG-TGTGCATTCTTGTTTCCTAA 539
           |||||||||||||||||||||||||||| ||||||||  ||||||||||||||||||||
Db     517 ATGTCACATGAGTGTGACTCCCTAAAGTAGAATTAAAGTTGTGCATTCTTGTTTCCCTA 576

Qy     540 AATATGA---TTTTACTGCTTGAAATT-ACATTTGAGTTG-AAGTTTAGAAAACTAACAT 593
           |||||||   |||||||||||||||||  |||||||||||  ||||||||||||||||||
Db     577 ACACATATGATTCTTACTGCTTGAAATTCACATTTGAGTTGAAAGTCTAGACACTTAACAT 636

Qy     594 AGCATTA-ATATGAATAATGCATGAAAAATTATATCCTTTGAAAACTGATTGATAAATA 652
           ||||||| |||||||||||||||||||||||||||||||||||  ||||||||||||||
Db     637 AGCATTACATATGAATAATGCATGGCACAATACTCATCTTTGCACACCGGTATGACACTT 696

Qy     653 TATTCCCCTCCTTTAGAAAC 673
           |||||||||||||| ||||||
Db     697 TTTTCCCCTCCTTTAAAACAC 717
```

RESULT 10
AV719129/c
LOCUS   AV719129   589 bp   mRNA   linear   EST 16-OCT-2000
DEFINITION   AV719129 GLC Homo sapiens cDNA clone GLCFLG09 5', mRNA sequence.
ACCESSION   AV719129
VERSION   AV719129.1   GI:10816281
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 589)
AUTHORS   Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z.,
              Xiao,H., Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z.,
              Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M.,
              Lu,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
TITLE   Homo sapiens cDNA GLC clones
JOURNAL   Unpublished (2000)
COMMENT   Contact: Zeguang Han
              Chinese National Human Genome Center at Shanghai
              351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
              201203, P. R. China
              Tel: 86-21-50801919(ex.45)
              Fax: 86-21-50801922
              Email: hanzg@chgc.sh.cn
              This clone is available at CHGC in Shanghai.
FEATURES   Location/Qualifiers
     source   1..589
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GLCFLG09"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="GLC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

Query Match   30.0%;   Score 486;   DB 9;   Length 589;

; Mon Apr 26 09:48:38 2004

```
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL.; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 345.
```

FEATURES
source   Location/Qualifiers
             1..618
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="GDB:5217005"
             /db_xref="taxon:9606"
             /clone="IMAGE:646753"
             /dev_stage="Ntera-2/RA neuroepithelial cells"
             /lab_host="SOLR (kanamycin resistant)"
             /clone_lib="Stratagene neuroepithelium (#937231)"
             /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
             XhoI; Cloned unidirectionally. Primer: Oligo dT. NT̄2
             cells (Ntera-2/c1.D1) induced with Retinoic Acid for 24
             hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; ~5'
             adaptor sequence: 5' GAATTCGGCACGAG 3' .-3' adaptor
             sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match   29.5%;   Score 477.8; DB 9;   Length 618;
Best Local Similarity 93.6%;   Pred. No. 8.3e-87;
Matches 573;   Conservative 0; Mismatches 30;   Indels 9;   Gaps 7;

Qy   1008 AACCATATATTGAATCTCTGATAAATATTTTCTGATGTACTAGCTATGGGAAATTAGAAC 1067
            ||||| ||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   603 AACCCATATGGAACNCGGATAAATATTTTTCCGGANGTACTNGCTATGGGAAATTGGAAC 544

Qy   1068 TGGCACAACCCTGACATTACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGA 1127
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   543 TGGCACACCCCTGACATTACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGA 484

Qy   1128 ATCTCTAAAATTTAAACATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATCAATGAATT 1187
            ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   483 ATCTCTAAAATTTAAACATT CCTGTTAAATGACTAAGGTTTGCTTTTATCAATCAATGAATT 424

Qy   1188 CTGAAGGCCAATATCATACCATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTTAG 1247
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   423 CTGAAGGCCAATATCATACCATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTTAG 364

Qy   1248 AGATATTCAAGCAATGCTCTCCTAATATCCATACGCAAGTGTGTTTATGACACAAAATTCA 1307
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   363 AGATATTCAAGCAATGCTCTCCTAATATCCATACGCAAGTGTGTTTATGACACAAAATTCA 304

Qy   1308 CTAGTCTGGTTTAAAAATGAAATCTTATATTGACTGGGGTGTCCCACATATTCCAGTAAT 1367
            |||||| |||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   303 CTAGTCT-GTTTAAAAATGAAATCTTATATTGACT-GGTGTCCCACATATTTCAGTAAT 246

Qy   1368 TTCTGTTATGAGAGACTTGAAATGCCACACAGTTAACTGATAGATCACG 1427
            |||||||||||||||||||| |||||||||||||||||||||||||||
Db   245 TTCTGTTATGAGAGACTTGAAATAGCAAATTG-CCACACAGTTAACTGATAATGATGATGACG 187

Qy   1428 TACGTGGTGATCATAACCACTTGGTACTACACCCCAGAAACTCAAAATGTCTTTTCTCCT 1487
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   186 TACGTGGTGATCATAACCACTTGGTACTACACCCCAGAAACTCAAAATGTC-TTTCTCCT 128

Qy   1488 GATGAGATATGGGGTGTCCTTTGTACGTGTCTAGGGCTACCCAAGTGAAGTGAAT 1547
            ||||||||| ||||||||||||||||||||||||||
Db   127 GATGAGATAT-GGGTGTCCTTTGTACGTGTCTAGGGCC---TAGGTAACCAGTGAGTGAATT 72

Qy   1548 ATATAAGCAAAATGTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTT 1607
            |||| ||||||| ||||| ||||||||||||||||||||||||||||||||||||
Db   71   ATATTAGC-AAATGTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTT 13

Qy   1608 AAAAATTTAAAA 1619
            ||||||||||||
Db   12   AAAAATTTAAAA 1

RESULT 12
```

Z98443/c
LOCUS   Z98443   580 bp   mRNA   linear   EST 22-SEP-1997
DEFINITION   HSZ98443 DKFZphthml Homo sapiens cDNA clone DKFZphthml_1e10 3',
             mRNA sequence.
ACCESSION   Z98443
VERSION   Z98443.1   GI:2326591
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 580)
AUTHORS   Korn,B., Wiemann,S., Ebert,L. and Poustka,A.
TITLE   Human ESTs (Korn,B. et al.)
JOURNAL   Unpublished (1997)
COMMENT   Contact: B. Korn
             Institution Molekulare Genomanalyse
             Deutsches Krebsforschungszentrum Heidelberg
             Im Neuenheimer Feld 506, D-69120 Heidelberg, FRG.
FEATURES   Location/Qualifiers
source   1..580
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="DKFZphthml_1e10"
                /tissue_type="thymus"
                /dev_stage="adult"
                /clone_lib="DKFZphthml"

ORIGIN

Query Match   29.3%;   Score 475;   DB 14;   Length 580;
Best Local Similarity 95.7%;   Pred. No. 3.1e-86;
Matches 563;   Conservative 0; Mismatches 16;   Indels 9;   Gaps 7;

Qy   1025 TGATAAAATATTTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACACCCTGACAT 1084
            ||||| || |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   580 TGATANATATTTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACACCCCTGACAT 521

Qy   1085 TACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAAATCTCTAAAATTTAAAC 1144
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   520 TACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAAATCTCTAAAATTTAAAC 461

Qy   1145 ATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCAT 1204
            ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   460 ATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATGAATTCTGAAGGCCAATATCAT 401

Qy   1205 ACCATTAACTATGAAAGCTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCAATGC 1264
            ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   400 ACCATTAACTATGAAAGCTTTATTCCTAAAAATAGTTTTAGAGATATTCAAGCAATGC 341

Qy   1265 TCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCTCGTGTTAAAAA 1324
            |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   340 TCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCT-GTTTAAAAA 282

Qy   1325 TGAAATCTTATATTGACTGGGGTGTCCCACATATTCCGTATGATGAGGAC 1384
            |||||||||||||||||||| |||||||||||||||
Db   281 TGAATTCTTATATTGACT-GGTGTTCCACATATTTCGTTATGAGAGGAC 223

Qy   1385 TTGAAAATAGCAAATTGCCCAACAGTTAACTGATGATCACTGGTACTACATAAC 1444
            |||| |||||||||||||||||||||||||||||||||||||| ||||||
Db   222 TTGAAATAGCAAATTG-CCACACAGTTAACTGATGACCACTGGTGGTCATAAC 164

Qy   1445 CACTTGGTACTACACCCAGAAACTCAAAATGTCTTTTCTCCGATGAGATATGGGGTGT 1504
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   163 CACTTGGTACTACACCCAGAAACTCAAAATGTC-TTTCTCCTGATGAGATAT-GGGTGT 106

Qy   1505 CCTTTTGTACGTCTAGGGCCTAGGGTACCCAAGTGAAGTGAATATATAAGCAAAATGTGT 1564
            ||||||||||||| ||||||| |||||||||||||||
Db   105 CCTTTTGTACGTCTAGGCC---TAGGTAACCAGTGAGTGGATTATTAGC-AAATGTGT 50

Qy   1565 TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAA 1612
            |||||||||||||||||||||||||||||| |||| ||||||||||||
Db   49   TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAA 2

RESULT 13

```
AW969587
LOCUS       AW969587            697 bp    mRNA    linear    EST 01-JUN-2000
DEFINITION  EST381664 MAGE resequences, MAGK Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW969587
VERSION     AW969587.1  GI:8159431
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C.,
            Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 267
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1..697
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGK"
                     /note="Vector: pBluescriptSKm"
ORIGIN
```

```
Query Match       29.3%;  Score 474.2;  DB 10;  Length 697;
Best Local Similarity  95.4%;  Pred. No. 4.4e-86;
Matches  564;  Conservative  0;  Mismatches  18;  Indels  9;  Gaps  7;

Qy  1028  TAAATATTTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACATTAC  1087
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  TAAATATTTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACATTAC    60

Qy  1088  TAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCAAAATTTAAACATT  1147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  TAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCAAAATTTAAACATT   120

Qy  1148  CCTGTTAAATGACTAAGGTTTGCTTTTATCAATATGAATTCGAAGGCCAATATCATACC  1207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CCTGTTAAATGACTAAGGTTTGCTTTTATCAATATGAATTCGAAGGCCAATATCATACC   180

Qy  1208  ATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCAATGCTCT  1267
          |||||||||||||||||||||||  |||||||||||||||||||||
Db   181  ATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCAATGCTCT   240

Qy  1268  CCTAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCTCGGTTTAAAAATGA  1327
          |||||||||||||||||||||  ||||||||||||||||  |||||||||||
Db   241  CCTAATATCCATACGCAAGTGTTTTATGACACAACAAATTCACTAGTCT-GTTTAAAAATGA   299

Qy  1328  AATCTTTATATTGACTGGGTGTCCCACATATTCCAGTAATTTCTGTTATGAGGACTTG  1387
          |||||||||||||||||||  ||||||||||||||||||  |||||||||||||
Db   300  AATCTTTATATTGACT-GTGTTCCACACATATTCAGTAATTTCTGTTATGAGGACTTG   358

Qy  1388  AAATAGCAAATTGCCCACACAGTTAACTGGATAGATCACGTACGTGGTCATAACCAC  1447
          ||||||||||||||||||||  ||||||||  |||||||||||||||||||||
Db   359  AAATAGCAAATTG-CCACACAGTTAACTGGATAGACCACGTACGTGGTCATAACCAC   417

Qy  1448  TTGGTACTACACCCAGAAACTCAAAATTGTCTTTCTCCTGATGAGGATATGGGGTGTCCT  1507
          ||||||||||||||||||||||||||  ||||  |||||||||||||||||||
Db   418  TTGGTACTACACCCAGAAACTCAAAATTGTC-TTTCTCCTGATGAGGATATAT-GGGTGTCCT   475
```

RESULT 14

```
BE564550
LOCUS       BE564550            739 bp    mRNA    linear    EST 15-AUG-2000
DEFINITION  601343885F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3686108 5',
            mRNA sequence.
ACCESSION   BE564550
VERSION     BE564550.1  GI:9808270
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 739)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM375  row: i  column: 21
            High quality sequence stop: 534.
FEATURES             Location/Qualifiers
     source          1..739
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3686108"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_53"
                     /note="Organ: bladder; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI
                     (ggccattatggcc); Double-stranded cDNA was prepared from
                     cell line RNA. 5' and 3' adaptors were used in cloning as
                     follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and
                     3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.55
                     kb (range 0.9-4.0 kb). 15/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA)."
ORIGIN
```

```
Query Match       28.7%;  Score 464.4;  DB 10;  Length 739;
Best Local Similarity  84.5%;  Pred. No. 4.2e-84;
Matches  607;  Conservative  0;  Mismatches  96;  Indels  15;  Gaps  7;

Qy  1508  TTTGTACGTCTAGGGCCTAGGTACCCAAGTGAGTGAATATATAAGCAAAATGTGTTTG  1567
          ||||||||||||||    ||||||||||||||||||||||||||||||||
Db   476  TTTGTACGTCTAGGCC---TAGGTAACCAGTGAGTGATTATATTAGC-AAATGTGTTTG   531

Qy  1568  TATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTATTTAAAAATTTAA  1618
          ||||||||||||||||||||||||||||||||  ||||||||
Db   532  TATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTATTTAAAAAAAAAAA   582

Qy   368  TAGACAGGAGATGGGGAAATTAGAAAGAGAACTGTGTAATTGAAATGAGTGGGCTGCA  427
          ||||||||||||||  ||||||||||||||||||||||||||||||||||
Db     1  TAGACAGGAGATGGGGAAATTAGAAAGAGAACTGTGTAATTGAAATGAGCGGGCTGCA   60

Qy   428  CCCTTAAGGAACTTATAATTAATAATGATGATCTGAATAAACATACCAGGATAAAGATGTCAA  487
          ||||||||||||||||||||||||||||||||  ||||||||||||||||
Db    61  CCCTTAAGGAACTTATAATTAATAATGATGATCTGAATAAACATACCAGGATAAAGATGTCAA   120

Qy   488  ATGAGTGTGACTCCCTTAAAGTAAGTAGATTAAAGTGTGCATTCTTGTTCTTGTTGTTCCTAAAAATATGAT  547
          |||||||||||||||||||||||||||||||||||||||||  ||||||||
Db   121  ATGAGTGTGACTCCCTTAAAGTAAGTAGATTAAAGTGTGCATTCTTTGTTCTTGTTTTCCTAAAAATATGAT   180
```

```
Query Match           28.5%;  Score 461.2;  DB 10;  Length 626;
Best Local Similarity 94.5%;  Pred. No. 1.9e-83;
Matches 554;  Conservative 0;  Mismatches 23;  Indels 9;  Gaps 7;

Qy  1028 TAAATATTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACATTAC 1087
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 TAAATATTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACATTAC   60

Qy  1088 TAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCTAAAATTTAAACATT 1147
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCTAAAATTTAAACATT  120

Qy  1148 CCTGTTAAATGACTAAGGTTTGCTTTTATCAATAGAATTCTGAAGGCCAATATCATACC 1207
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCTGTTAAATGACTAAGGTTTGCTTTTATCAATAGAATTCTGAAGGCCAATATCATACC  180

Qy  1208 ATTAACTATGAAAGCTTTTAATTCCTAAAATAGTTTTAGAGATATTCAAGCAATGCTCT 1267
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ATTAACTATGAAAGCTTTTAATTCCTAAAATAGTTTTAGAGATATTCAAGCAATGCTCT  240

Qy  1268 CCTAATATCCATACGCAAGTGTGTTTATGACACAAATTCACTAGTCTGGTTTAAAATGA 1327
            |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   241 CCTAATATCCATACGCAAGTGTGTTTATGACACAAATTCACTAGTCT-GTTAAAATGA  299

Qy  1328 AATCTTTATATTGACTGGGTCCCACATATTCCAGTAATTCTGTTATGATGAGGACTTG 1387
            |||||||||||||| |||  |||||||| ||||||||||||||||||||||||||||
Db   300 ATTCTTTATATTGACT-GGTGTTCCACATATTTCAGTAATTCTGTTATGATGAGGACTTG  358

Qy  1388 AAATAGCAAATGTGCCCACACAGTAACTGGATAGATCACGTACGTGGTGATCATAACCAC 1447
            |||||||||  |  |||||||||||||||||||| ||||||||||||||||||||||||
Db   359 AAATAGCAAATTG-CCACACAGTAACTGGATAGACCACGTACGTGGTGATCATAACCAC  417

Qy  1448 TTGGTACTACACCCAGAAACTCAAAATTGTCTTTTCTCCTGATGAGATATGGGGTGTCCT 1507
            ||||||||||||||||||||||||||||||| ||||||||||||||| |||||||||||
Db   418 TTGGTACTACACCCAGAAACTCAAAATTGTC-TTTCTCCTGATGAGATAT-GGGTGTCCT  475

Qy  1508 TTTGTACGTCTCAGGGCTCTAGGGCC---TAGTAACCAGTGGAGTGAATATATAAGCAAAATGTGTTTG 1567
            ||||||||||||||||||   |||       ||||||||||||||||| |||||| || |||||||||||
Db   476 TTTGTACGTCTCAGGGCC---------TAGTAACCAGTGGAGTGATTATATTAGC-AAATGTGTTTG  531

Qy  1568 TATCCAGAGTCTTCCTGTCATTGTAATAAAAATTTATTTAAAAAT 1613
            ||||||  ||||  ||| |||||||||||||||| |||||||||
Db   532 TATCCAAAGTCTCCTGACATTGTAATAAAAATTTTATTTAAAATT 577
```

Search completed: April 26, 2004, 01:59:22
Job time : 2943.54 secs

```
Qy   548 TTTACT-GCTTGAAAATTACATTTGAGTTGAAGTTTAGAAACATAGCATTAATATGA 606
            |||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TTTACTGCTTGAAAATTACATTTGAGTTGAAGTTTAGAAACATAGCATTAATATGA 240

Qy   607 ATAATGCATGGAAAATTATTATCCTTTGAAAACTGATTGATAAATATTCCCCCTCCTT 666
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ATAATGCATGGAAAATTATTATCCTTTGAAAACTGATTGATAAATATTCCCCCTCCTT 300

Qy   667 TAGAAACAGTCAAAAGCCACTTCAAACAGTTCAAAATAAAGAAGGTAGCAAGTTAGG 726
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TAGAAACAGTCAAAAGCCACTTCAAACAGTTCAAAATAAAGAAGGTAGCAAGTTAGG 360

Qy   727 CGATGGATTATATTTTCTTGGCTTGTGTATACCCATTGGCCAGGGCCTTTATAAGGACT 786
            |||||||||||||||||||  ||||||||||||  |||  ||||  |||||||||||||
Db   361 CGATGGATTATATTTTCTTG--CTTGTGTATACCCAGTTGTCAAGGACATTATAAGGACTC 418

Qy   787 CCCAAAGCATTTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCAAATGAGGCA 846
            ||||||||||| |||| ||||||||||||||||||||||||||||  ||||||||||
Db   419 CCAAAGCATTTCGAAGGATGGCAATATCAAATAAGTGTATGTCCTCTCAAATGAGGCA 478

Qy   847 TTTTTAATTGTTACAA--TCTATTTGGACGGCTCAGGTT---ATGATATGTTTATGAAAA 900
            ||||||| |||  ||  |||| ||||||| |||||||||   |||||||||||||||| |
Db   479 TTTTTAAATTGGTTAAAAATCTAATTGGACGCTCAGGTTTATGATATGTTTATGAAACAA 538

Qy   901 ATAAGCTTCATTATTTCTTATAGCTACATCCTATTATT--CCCTTTTAGAAACAAGAA-T 957
            |||||| ||||||||| |||||| || ||  ||||||| |||||| ||||||||||| |
Db   539 TAAGCCTTCATTATTTTTATAGGCTAAATTTCATTATTTCCCCTTTGGAAACAAGAATT 598

Qy   958 AACAATAAGTTTTATAATGTTGCCATACTTAGCATTATCAGTTCTAATGAAACCAATATT 1017
            ||| ||| ||||| || ||| | ||||||||||||||||  ||||| ||||| |||| ||
Db   599 ACCATTAGGTTTTAATAGTGCATAACTTAGCATTATCAGGGTCTAAGGAACACAATTT 658

Qy  1018 GAATCTCT--GATAAATATTTCTGATGTTACTAGCTATGGGAAAATTAGAACTGGCA 1072
            |||||||| |  |||| |||||| ||| ||| || |||||||  |||||| ||||||
Db   659 TGGAATCTCTGGGGTAAAAATTTTCGGGTGGGTACAGACCTTGGGGAATTAGAAATGGGGCA 716
```

RESULT 15
AW969584

```
LOCUS       AW969584     626 bp    mRNA    linear    EST 01-JUN-2000
DEFINITION  EST381661 MAGE resequences, MAGK Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW969584
VERSION     AW969584.1  GI:8159428
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 626)
AUTHORS     Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C.,
            Holt,I.B., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
TITLE       Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
JOURNAL     Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 267
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1..626
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGK"
                     /note="vector: pBluescriptSKm"
ORIGIN
```

OM nucleic - nucleic search, using sw model

```
          Copyright  GenCore version 5.1.6
                     (c) 1993 - 2004 Compugen Ltd.
```

Run on: April 25, 2004, 07:15:18 ; Search time 80.3165 Seconds
                                       (without alignments)
                                       11186.557 Million cell updates/sec

Title: US-10-051-835-15
Perfect score: 1619
Sequence: 1 ttaatctacatatccttagt........atttattaaaatttaaaa 1619

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : Issued Patents_NA:*
            1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
            2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
            3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
            4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
            5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	59.8	3.7	7218	1	US-08-232-463-14	Sequence 14, Appl
c	2	47.2	2.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	3	46.4	2.9	399	4	US-09-621-976-8976	Sequence 8976, Ap
	4	45.4	2.8	11049	4	US-10-204-708-22	Sequence 22, Appl
	5	44.6	2.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	6	44.4	2.7	640681	4	US-09-790-988-1	Sequence 1, Appli
c	7	44.4	2.7	640681	4	US-09-790-988-1	Sequence 1, Appli
c	8	43.6	2.7	8654	1	US-08-920-812-6	Sequence 6, Appli
c	9	43.6	2.7	8654	1	US-08-920-827-6	Sequence 6, Appli
c	10	43.6	2.7	8654	1	US-08-921-177-6	Sequence 6, Appli
c	11	43.6	2.7	8654	1	US-08-362-577C-6	Sequence 6, Appli
c	12	43.6	2.7	8654	2	US-08-920-828-6	Sequence 6, Appli
	13	43.2	2.7	832	4	US-09-621-976-2813	Sequence 2813, Ap
c	14	43.2	2.7	1542	4	US-09-202-893B-3	Sequence 3, Appli
c	15	42.8	2.6	832	4	US-09-621-976-2813	Sequence 2813, Ap
c	16	42.2	2.6	493	4	US-09-919-172-35	Sequence 35, Appl
c	17	42.2	2.6	1500	4	US-09-601-198-36	Sequence 36, Appl
c	18	40.8	2.5	6156	4	US-10-204-708-60	Sequence 60, Appl
	19	40.8	2.5	7198	4	US-09-453-702B-147	Sequence 147, App
c	20	40.6	2.5	64467	4	US-09-803-671B-13	Sequence 13, Appl
	21	40	2.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
	22	39.8	2.5	2861	1	US-08-299-953-1	Sequence 1, Appli
	23	39.8	2.5	2861	1	US-08-459-415-1	Sequence 1, Appli
	24	39.8	2.5	2861	1	US-09-066-687-1	Sequence 1, Appli
	25	39.8	2.5	3881	5	PCT-US95-11231-1	Sequence 1, Appli
	26	39.8	2.5	3881	1	US-08-299-953-2	Sequence 2, Appli
	27	39.8	2.5	3881	1	US-08-459-415-2	Sequence 2, Appli
c	28	39.8	2.5	3881		US-09-066-687-2	Sequence 2, Appli
c	29	39.8	2.5	3881	5	PCT-US95-11231-2	Sequence 2, Appli
	30	39.6	2.4	2614	4	US-09-004-056-1	Sequence 1, Appli
	31	39.4	2.4	19124	2	US-08-487-826B-13	Sequence 13, Appl
	32	39.2	2.4	319608	4	US-09-539-333D-1	Sequence 1, Appli
	33	39.2	2.4	319608	4	US-09-679-409-1	Sequence 1, Appli
	34	39	2.4	2169	4	US-09-434-408-3	Sequence 3, Appl
c	35	39	2.4	162450	4	US-09-345-882-1	Sequence 1, Appli
	36	38.8	2.4	1189	1	US-08-307-591-2	Sequence 2, Appli
c	37	38.8	2.4	53332	4	US-09-801-861-3	Sequence 3, Appl
	38	38.6	2.4	665	2	US-08-883-795A-36	Sequence 36, Appl
	39	38.6	2.4	9347	4	US-10-204-708-35	Sequence 35, Appl
c	40	38.2	2.4	665	2	US-08-883-795A-36	Sequence 36, Appl
c	41	38.2	2.4	1149	4	US-09-601-198-151	Sequence 151, App
c	42	38.2	2.4	10607	1	US-08-078-090-3	Sequence 3, Appli
	43	38	2.3	2232	4	US-08-956-171E-210	Sequence 210, App
	44	38	2.3	3400	4	US-09-694-127-1	Sequence 1, Appli
	45	38	2.3	168575	4	US-09-426-290-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;  APPLICANT: DORNER, F.
;  APPLICANT: SCHEIFLINGER, F.
;  APPLICANT: FALKNER, F. G.
;  TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;  NUMBER OF SEQUENCES: 52
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: Foley & Lardner
;  STREET: 1800 Diagonal Road, Suite 500
;  CITY: Alexandria
;  STATE: VA
;  COUNTRY: USA
;  ZIP: 22313-0299
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/232,463
;  FILING DATE:
;  CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: US/07/935,313
;  FILING DATE:
;  APPLICATION NUMBER: EP 91 114 300.6
;  FILING DATE: 26-AUG-1991
;  ATTORNEY/AGENT INFORMATION:
;  NAME: BENT, Stephen A.
;  REGISTRATION NUMBER: 29,768
;  REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (703)836-9300
;  TELEFAX: (703)683-4109
;  TELEX: 899149
;  INFORMATION FOR SEQ ID NO: 14:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 7218 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  IMMEDIATE SOURCE:
;  CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match             3.7%;  Score 59.8;  DB 1;  Length 7218;
```

```
Best Local Similarity  2.6%; Pred. No. 5.5e-05;
Matches   10; Conservative 214; Mismatches 155; Indels   0; Gaps   0;

Qy     21 ATATAAATATTGCCTAATCCATAGAAAAAGGATATAAAGTATTAAATATGTGATATATA 80
             |||  ||  |||
Db   1451 ATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1392

Qy     81 GCNNNNNNNNNNNNNNNNNNNNNNNATAGGGAAGTTCAAGTCACTTCAATTGAAGAAAC 140
                            :: ::::::
Db   1391 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1332

Qy    141 ATATCTCTGAGCATAGGAGCAGCCTCAGGTCCTATGGTGGCAGTGACGTGGACAGGAGAGG 200
                  :: : ::::: :::
Db   1331 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1272

Qy    201 GGGAAATTAGAAAAGAGAAACTATATATATTGAAAAAGGGATATAAAGCATTAAATATATGA 260
            ::::::::::::
Db   1271 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1212

Qy    261 TATATAGCTATATCTATGTATGTATCTAACAGAGAAGTTCAAGTCACTTCCAATTAAAGAA 320
                ::: :: ::::: ::::
Db   1211 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1152

Qy    321 ACATTTTGAGCATGGGACCAGCCCTCAGGTCTTATGCTGGGATGCAGTAGACAGGAGATG 380
               ::: :: ::::: ::: : :::::
Db   1151 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1092

Qy    381 GGGAAATTAGAAAAGAGAA 399
            ::::: :::::::::
Db   1091 RRRRRRRRRRRRRRRRRRR 1073


RESULT 2
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc
; Patent No. 6503729
; FILE REFERENCE: PB275
; TITLE OF INVENTION: jannaschii
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 1
;   LENGTH: 1664976
;   TYPE: DNA
;   ORGANISM: Methanococcus jannaschii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (28222)..(28222)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (28257)..(28258)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84773)..(84773)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84808)..(84808)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (84812)..(84812)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98120)..(98120)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98159)..(98159)
```

```
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98239)..(98239)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98266)..(98266).
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (98343)..(98343)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (103998)..(103998)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (148948)..(148948)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (163385)..(163385)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191989)..(191989)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (191995)..(191995)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (231980)..(231980)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234187)..(234187)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234220)..(234220)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (234814)..(234814)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309398)..(309398)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (309418)..(309418)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312837)..(312837)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (312993)..(312993)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (319226)..(319226)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559167)..(559167)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (559241)..(559241)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (600992)..(600992)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (622708)..(622708)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657081)..(657081)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (657203)..(657203)
;   OTHER INFORMATION: n equals a, t, c, or g
;   NAME/KEY: misc_feature
;   LOCATION: (674435)..(674435)
;   OTHER INFORMATION: n equals a, t, c, or g
```

NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664855)..(1664855)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match   2.9%; Score 47.2; DB 4; Length 1664976;
Best Local Similarity 52.6%; Pred. No. 0.3;
Matches 103; Conservative   0; Mismatches 93; Indels   0; Gaps   0;

Qy   838 AATGAGGCATTTTTAATTGTTACAATCTATTTGGACGCTCAGGTTATGATATGTTTATGA 897
               ||| |||||||| ||||| ||| |||| |||||||||| |||| |||||||||
Db   317507 AATAACGCATTTTCTAATGGTTCGATGTCTTTTACCTTTCTTGATAATTTTAACCATA 317448

Qy   898 AAAATAAGCTTCATTATTTCTTATAGCTACATCCTATTATTCCCTTTAGAAACAAGAAT 957
               ||| ||||||||||| |||| ||||| ||||||||||||||||||| |||||||||
Db   317447 AACATCACCTTAATATTTATTATCATTAGAGCTTTTATATTTTTGCAAACTCCCTT 317388

Qy   958 AACAATAAGTTTTAATAGTTGCCATACTTAGCATTTATCAGGTCTAATGAAACCAATATT 1017
               || |||||||| |||| |||| ||| | ||||||| ||||||||||||||||||||
Db   317387 ATAAATTTTATTTCATCAATTCCAAATATTTAGATTTTCTCCTTATGGTTTTTGCTATTCTT 317328

Qy   1018 GAATCTCTGATAAATA 1033
               ||||||| ||||||
Db   317327 CCTTCTCCAAACAATA 317312

RESULT 3
US-09-621-976-8976
; Sequence 8976, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8976
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-8976

Query Match   2.9%; Score 46.4; DB 4; Length 399;
Best Local Similarity 10.1%; Pred. No. 0.037;
Matches 37; Conservative 182; Mismatches 143; Indels   4; Gaps   1;

Qy   464 AACATACCAGGATAAAGATGTCAAATGAGTGTGACTCCCTTAAAGTAGATTAAAGTGTGC 523
               :: |:||:||:||:||: |: :|| :| |:: |:: ::   :: :| ::
Db   1 RWRAYMCWAKGAKWARAMTRATWGMYTRGSKKSWRAKMSWMKKRRRRRARWMWKSMCWM 60

Qy   524 ATTCTTTGTTCCTAAAATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAGTTTAG 583
               :: :|:||: ::: ||| :|: :: |:: |:|::|: ||::|||:|:||::|::|:
Db   61 KKSKSWWRSWGMTKRMKGRGAASWAGYMSWMTYMTRRWRYRYRKKACTKWRAAGMWGK 120

Qy   584 AACTAACATAGCATTAATATGAATAATGCATGGAAAATTATTATCCTTGAAAACTGAT 643
               |:| :|:||||:| :||:|:|: :|::||| :||:|:::|:| :||||::|||
Db   121 AGWAWMAYAKMWTMAWRRTAMKYWAMMKSKRSMRRRRAWYAW---MYYMMARRTMWGMR 176

Qy   644 TGATAAATATATTCCCCTCCTTTAGAACAGTCAAAAGCCACTTCAAACAAGTTTCAAA 703
               :::: : :|: :::| ::|:::||:| :| :: : ::   ::|::| :
Db   177 ASCYRGAYMASAGMYYWMMYYMMRRKWMYSAGWSMMRKWTRRCASYSCWSSYCMWGAXWM 236

Qy   704 ATAAAGGAAGGTAGCAAGTTAGGCGATGGATTATATTTCTTGGCTTGTTGTATACCCAT 763
               ||| ::|: ::|: |:|:|::||::|: ||:|:|:::||:::|: ||:|:::|:|:::
Db   237 MYWKTSRWSYWYSSYRCTKYRRSCCCWSMSCYWKTYYRSWYCASCYYSYKTKRASCCMMC 296

Qy   764 TGGCCAGGGCCTTTATAAGGACTCCCAAAAGCATTTGAAGAATGGCAATATCAAATAAG 823
               |||||||| ::::|:|||::|::|:||||| |||:|||| :|| |:|||:|:|||:||
Db   297 CCMKWRKMKMAMMWMKMTYCKKTSAMRYRSCYSAKRRWRMWCCAYWMKTCSWWMMCWT 356

Qy   824 TGTATG 829
               |||:||
Db   357 TGGCTG 362

RESULT 4
US-10-204-708-22
; Sequence 22, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
TITLE OF INVENTION: by Assessing DNA Methylation
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204,708
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98

SEQ ID NO 22
LENGTH: 11049
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-22

Query Match   2.8%;   Score 45.4; DB 4;   Length 11049;
Best Local Similarity   48.3%;   Pred. No. 0.18;
Matches 156;   Conservative   0;   Mismatches 166; Indels   1;   Gaps   1;

Qy   796 ATTTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCTCAAATGAGGCATTTTAATT 855
          ||||| |||||| | |||| | |||| | ||||| | | ||||||||||
Db   254 ATTGTAAAGAATTGTTTTTTTAAAAAAAAATAAGTGATTGTTAAAGAATTTTATGAATT 313

Qy   856 GTTACAATCTATTTGGACGCTCAGGTTATGATGTTTATGAAAATAAGCTTCATTATT 915
          | ||||| |||| ||||| ||||| | | |||| |||||| | ||| ||||||
Db   314 AGTATAAATTTTTTGGAGGGATAAATTTGATAAT-TTATAGTGAAAAAAATTTTTAAAATT 372

Qy   916 TCTTATAGCTACACATCCTATTATTCCCTTTAGAAACAAGAATAACAATAAGTTTTAATAG 975
          | || || | | || || | ||||| | ||| ||| |||| |||||||
Db   373 TTGTAGTTTATTTTATTATTTTTAGAATATTGATTTTTTAGAAAAATTATTATGATATATAA 432

Qy   976 TTGCCATACTTAGCATTTATCAGGTCTAATGAAACCAATATTGAAATCTCTGATAAATATT 1035
          ||| || ||||| | ||| |||| | ||| ||||| ||||| || ||| |||||
Db   433 TGAGAGTTATTTATAGAATGTTTGTT-ATAATAATGAAAATATAAAAATTAATTTAAAAG 492

Qy   1036 TTCTGATGTTACTAGCTATGGGAAAATTAGACTGGCACAACCCTGACATTACTAAGTGGA 1095
          ||| || |||| ||||||| | |||| | ||| || ||| |||| ||| ||
Db   493 TTTTTTAGGTAGGTTTTTTAAAAAATTAATTATATAGGTCGGGCGGCCGGTGGTTTATATTTGT 552

Qy   1096 AATGTTAGGATTTTTCGGCATCG 1118
          ||| || ||| |||||||| |||
Db   553 AATTTTAGTATTTTGGGAGGTCG 575

RESULT 5
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1

SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA

ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103398)..(103398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)

```
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (559167)..(559167)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (559241)..(559241)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (600992)..(600992)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (622708)..(622708)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (657081)..(657081)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (657203)..(657203)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (674435)..(674435)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (682442)..(682442)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (713652)..(713652)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (741684)..(741684)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (779455)..(779455)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (779676)..(779676)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (855539)..(855539)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (871619)..(871619)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1084830)..(1084830)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1096846)..(1096846)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1119881)..(1119881)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1130881)..(1130881)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1310988)..(1310988)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1313224)..(1313224)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1349473)..(1349473)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1349491)..(1349491)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1470091)..(1470091)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1569020)..(1569020)
,;  OTHER INFORMATION: n equals a, t, c, or g
```

```
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1602912)..(1602912)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1603734)..(1603734)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1637998)..(1637998)
,;  OTHER INFORMATION: n equals a, t, c, or g
,;  NAME/KEY: misc_feature
,;  LOCATION: (1664854)..(1664855)
,;  OTHER INFORMATION: n equals a, t, c, or g
;  US-08-916-421B-1

Query Match    2.8%;  Score 44.6;  DB 4;  Length 1664976;
Best Local Similarity  47.3%;  Pred. No. 1.3;
Matches 134;  Conservative 0;  Mismatches 149;  Indels 0;  Gaps 0;

Qy     461 ATAAACATACCAGGATAAAGATGTCAAATGAGTGACTCCCTAAAGTAGATTAAAGTG    520
                                       |  ||||    |           ||||  ||||
Db  484933 AAAAAATTAACAAGATAAAAATGAACTAAAATAATTAATTATATTAATTAATTAGTTTT 484992

Qy     521 TGCATCCTTGTTCCTAAAATATGAATTTTACTGCTTGAAATTACATTTGAGTTGAAGTT   580
               |    |      ||       |||||
Db  484993 ACACATTTTCAGCATCGTCCGTAATATTTGAAACATATTATTCCAATATCACGTCATCTC 485052

Qy     581 TAGAAACTAACATAGCATTAATATGAATAATGCATGGAAAATTATTATCCTTGAAAACT   640
              || |||||||
Db  485053 TATAAGCTAACACACATTCTCAGCAACAAACCCTAATTTAAAATTAAAAATACTCATAT 485112

Qy     641 GATTGATAAAATATATTCCCCCTCCTTTAGAAACAGTCAAAAGCCACTTCAAACAAGTTC   700
              ||  ||
Db  485113 AATCCCCTAAAATTCATTAAGCTCTTTTAAAATAGATTCAAAAGGTACTTCCTCAAAATTTA 485172

Qy     701 AAAATAAAGGAAGGTAGCAAGTTAGGCGATGGATTGATTATATTTTC    743
Db  485173 CAATTACAACGTCATATTCTGGGGATTTTATCTTTATATTTTC 485215


RESULT 6
US-09-790-988-1
;  Sequence 1, Application US/09790988
;  Patent No. 6632935
;  GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT: WATANABE, HIDEMI
;  APPLICANT: HATTORI, MASAHIRA
;  APPLICANT: SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE: 2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
;  SEQ ID NO 1
;  LENGTH: 640681
;  TYPE: DNA
;  ORGANISM: Buchnera sp.
;  US-09-790-988-1

Query Match    2.7%;  Score 44.4;  DB 4;  Length 640681;
Best Local Similarity  46.2%;  Pred. No. 1.1;
Matches 147;  Conservative 0;  Mismatches 171;  Indels 0;  Gaps 0;

Qy     504 TAAAGTAGATTAAAGTGTGCATTCTTTGTTCCTAAAATATGATTTTACTGCTTGAAATT    563
              ||||      |||       |||||          ||||
Db  580418 TAAAAAATCTAGAAATTGATTTTTTATTATTAAAATTTTTACTTAACATT 580477

Qy     564 ACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTAATATGAATAATGCATGGAAAATT   623
Db  580478 GTATATGTTAAAAAATACTTAATACAATAAATAAAATAAAAATTTATCTTTATCAAAGATAG 580537
```

Qy   624   ATTATCCTTGAAAACTGATTGATAAATATATTCCCCCTCCTTTAGAAACAGTCAAAAGC 683
            ||||| ||||||||
Db 580538   ATAAAACAATAAATTGTTATAGTAGAATATATCTAAAAAGTTTATTGATTGAAAAAAAT 580597

Qy   684   CACTTCAAACAAGTTTCAAAATAAAGGAAGGTAGCAAGTTAGGCGATGGATTATATTTTC 743
            ||||| ||||
Db 580598   AATATTAAAAGAAAATAAAAATATTTTTACATTGAAAAACAGATCAGAAAAAATATATTA 580657

Qy   744   TTGGCTTGTTGTATACCCATTGGCCAGGGCCTTTATAAGGACTCCCAAAAGCATTTTGAA 803
            |||| |||
Db 580658   TTTTTTTATCTAAACCTAATTGGAAAATAATTAAAAAAATAATTAAAAAAAGAGTCTTCAAT 580717

Qy   804   GAATGGCAAATAATCAAATA 821
Db 580718   TCATTGATAATATAAAATA 580735

RESULT 7
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;   APPLICANT: SHIGENOBU, SHUJI
;   APPLICANT: WATANABE, HIDEMI
;   APPLICANT: HATTORI, MASAHIRA
;   APPLICANT: SAKAKI, YOSHIYUKI
;   TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;   FILE REFERENCE: 081356/0159
;   CURRENT APPLICATION NUMBER: US/09/790,988
;   CURRENT FILING DATE: 2001-02-23
;   PRIOR APPLICATION NUMBER: JP2000-107160
;   PRIOR FILING DATE: 2000-04-07
;   NUMBER OF SEQ ID NOS: 7
;   SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match   2.7%;   Score 44.4;   DB 4;   Length 640681;
Best Local Similarity   46.0%;   Pred. No. 1.1;
Matches   186;   Conservative   0;   Mismatches 216;   Indels   2;   Gaps   1;

Qy   564   ACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTAATATGAATAATGCATGGAAAATT 623
            ||||| ||||||| |||| ||| |||||||
Db   3155   ACATTTAATTTTCCATTAGAAACTCCAGTTTTTAAAAAAATCACAATAGATTTTAATCT 3096

Qy   624   ATTATCCTTGAAAACTGATTGATAAATATATTCCCCCTCCTTTAGAAACAGTCAAAAGC 683
            ||||| ||||||||||
Db   3095   TTATGAGATTGACGAGGCCATTGATAAATATACAATAGTTAATACCATAAAAATAAAAGCC 3036

Qy   684   CACTTCAAACAAGTTTCAAAATAAAGGAAGGTAGCAAGTTAGGCGATGGATTATATTTTC 743
            |||| ||||| |||||||||||||
Db   3035   TGTAGTGAAATTATTAAAATATGAAAATAGCCACGGTTAAAAAACATTGTGTC 2976

Qy   744   TTGGCTTGTTGTATACC--CATTGGCCAGGGCCTTTATAAGGACTCCCAAAAGCATTTTG 801
            |||| ||||||| ||||||||||||||||
Db   2975   CACCATGGCAGCAAACCTGCAATTAAATAAAAAATCATTTCACCTGCGTACATGTTTTCCA 2916

Qy   802   AAGAATGGCAATATCAAATAAAGTGTATGTCCCTCAAATGAGGCATTTTAATTGTTACA 861
            |||| ||||||| |||||
Db   2915   AATAATCGCAATCCCAAAGAAATGGGTTTCGAGACCAAATGACACAAATTCTAATATAAAA 2856

Qy   862   ATCCTATTTGGACGCTCAGGTTATGATATGTTTATGAAAAGCTTCATTATTTCTTAT 921
            ||||
Db   2855   TTAAAAATAGAAAATACAGGATGGTTGTAAAAGGTTGTAAAATAAGTTCTTTTAAAAAGCCT 2796

Qy   922   AGCTACATCCTATTATTCCCTTTAACAGTATAAAACAAGAATAACAATAA 965
            ||| |||||
Db   2795   ACATATCCTTAATTTTAACAGTATAAAAATAAAAATTAAAAAAAAA 2752

RESULT 8
US-08-920-812-6/c
; Sequence 6, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
;   APPLICANT: Ohno, Tsuneya
;   APPLICANT: Matsuhisa, Akio
;   APPLICANT: Uehara, Hirotsugu
;   APPLICANT: Eda, Soji
;   TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
;   NUMBER OF SEQUENCES: 25
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower, 233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/920,812
;   FILING DATE: 29-AUG-1997
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/362,577
;   FILING DATE: 27-MAR-1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Rin-Laure, Li-Hsien
;   REGISTRATION NUMBER: 33,547
;   REFERENCE/DOCKET NUMBER: 19036/32420
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312/474-6300
;   TELEFAX: 312/474-0448
;   TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 8654 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
;   ORIGINAL SOURCE:
;   ORGANISM: Staphylococcus epidermidis
;   STRAIN: Clinical Isolate SB-22
US-08-920-812-6

Query Match   2.7%;   Score 43.6;   DB 1;   Length 8654;
Best Local Similarity   50.0%;   Pred. No. 0.44;
Matches   109;   Conservative   0;   Mismatches 109;   Indels   0;   Gaps   0;

Qy   440   TTATAATTAATGATCTGAATAAACATACCAGGATAAAGATGTCAAATGAGTGTGACT 499
            |||| |||| |||||||
Db   2309   TTAAATTAAGGAAAATTAGAAACAATCTATTTTTTGAATATAGCAAATGCTACATTGC 2250

Qy   500   CCCTTAAAGTAGATTAAAGTGTCATTCATTTGTTCCTAAAATATGATTTTACTGCTTGA 559
            |||
Db   2249   TAATTCAAGTATAAAACTTAAAAGTAATAATTTTGTTTTATAAACACTTAATATGGAAA 2190

Qy   560   AATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTAAATGAATAATGCATGGAA 619
            |||
Db   2189   TTAATTTTTATTTAGTTATTAATTAACTAAAATAATATAGCATTAAAATATTCTTAAATTTA 2130

Qy   620   AATTATATCCTTGAAAACTGATTGATAAATATC 657
            ||||| |||
Db   2129   AATTAATAACTATATTATGACCGGACACTTTCGTATTTTC 2092

RESULT 9

US-08-920-827-6/c
; Sequence 6, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus epidermidis
; STRAIN: Clinical Isolate SE-22
US-08-920-827-6

Query Match   2.7%;   Score 43.6; DB 1; Length 8654;
Best Local Similarity 50.0%; Pred. No. 0.44; Indels   0; Gaps   0;
Matches 109; Conservative   0; Mismatches 109; Indels 109; Gaps   0;

Qy   440 TTATAATTAATGATGATCTGAATAAACATACCAGGATAAAGATGTCAAATGAGTGTGACT 499
              ||||| || || ||| || | | || | | ||||
Db   2309 TTAAATTAAGGAAAATTAGAGAAACAATCTTATTTTTTTGAATATAGCAAATGCTACATTGC 2250

Qy   500 CCCTTAAAGTAGATAAGTGTGCATTCTTTGTTTCCTAAAATATGATTTTACTGCTTGA 559
             ||| | | | || | | || | |||| | |
Db   2249 TAATTCAAGTATAAACTTTAAAAGATAATTGTTTATAAACACTTATTTGTTTACTTCT 2190

Qy   560 AATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTAATATGAATAATGCATGGAA 619
              | ||| | | || || ||| | || ||
Db   2189 TAATTTTATTTAGTTATAATTAACTAAATAATATAGCATTAAATAATATATTCTTAAATTTA 2130

Qy   620 AATTATTATCCTTTGAAAACTGATTGATAAATATATTC 657
             || || | ||
Db   2129 AATTAATACTATATTATGACGGACACTTTCGTATTTTTC 2092

RESULT 10
US-08-921-177-6/c
; Sequence 6, Application US/08921177

; Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus epidermidis
; STRAIN: Clinical Isolate SE-22
US-08-921-177-6

Query Match   2.7%;   Score 43.6; DB 1; Length 8654;
Best Local Similarity 50.0%; Pred. No. 0.44; Indels   0; Gaps   0;
Matches 109; Conservative   0; Mismatches 109; Indels 109; Gaps   0;

Qy   440 TTATAATTAATGATGATCTGAATAAACATACCAGGATAAAGATGTCAAATGAGTGTGACT 499
              ||||| || || ||| || | | || | | ||||
Db   2309 TTAAATTAAGGAAAATTAGAGAAACAATCTTATTTTTTTGAATATAGCAAATGCTACATTGC 2250

Qy   500 CCCTTAAAGTAGATAAGTGTGCATTCTTTGTTTCCTAAAATATGATTTTACTGCTTGA 559
             ||| | | | || | | || | |||| | |
Db   2249 TAATTCAAGTATAAACTTTAAAAGATAATTGTTTATAAACACTTATTTGTTTACTTCT 2190

Qy   560 AATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTAATATGAATAATGCATGGAA 619
              | ||| | | || || ||| | || ||
Db   2189 TAATTTTATTTAGTTATAATTAACTAAATAATATAGCATTAAATAATATATTCTTAAATTTA 2130

Qy   620 AATTATTATCCTTTGAAAACTGATTGATAAATATATTC 657
             || || | ||
Db   2129 AATTAATACTATATTATGACGGACACTTTCGTATTTTTC 2092

RESULT 11
US-08-362-577C-6/c
; Sequence 6, Application US/08362577C
; Patent No. 5807673
; GENERAL INFORMATION:

```
;       APPLICANT: Ohno, Tsuneya
;       APPLICANT: Matsuhisa, Akio
;       APPLICANT: Uehara, Hirotsugu
;       APPLICANT: Eda, Soji
;       TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
;       NUMBER OF SEQUENCES: 25
;       CORRESPONDENCE ADDRESS:
;          ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;          STREET: 6300 Sears Tower, 233 South Wacker Drive
;          CITY: Chicago
;          STATE: Illinois
;          COUNTRY: United States of America
;          ZIP: 60606-6402
;       COMPUTER READABLE FORM:
;          MEDIUM TYPE: Floppy disk
;          COMPUTER: IBM PC compatible
;          OPERATING SYSTEM: PC-DOS/MS-DOS
;          SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;          APPLICATION NUMBER: US/08/362,577C
;          FILING DATE: 27-MAR-1995
;          CLASSIFICATION: 536
;       ATTORNEY/AGENT INFORMATION:
;          NAME: Rin-Laures, Li-Hsien
;          REGISTRATION NUMBER: 33,547
;          REFERENCE/DOCKET NUMBER: 19036/32420
;       TELECOMMUNICATION INFORMATION:
;          TELEPHONE: 312/474-6300
;          TELEFAX: 312/474-0448
;          TELEX: 25-3856
;       INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;          LENGTH: 8654 base pairs
;          TYPE: nucleic acid
;          STRANDEDNESS: double
;          TOPOLOGY: linear
;       MOLECULE TYPE: Genomic DNA
;       ORIGINAL SOURCE:
;          ORGANISM: Staphylococcus epidermidis
;          STRAIN: Clinical Isolate SE-22
US-08-362-577C-6

Query Match         2.7%;    Score 43.6; DB 1; Length 8654;
Best Local Similarity   50.0%;  Pred. No. 0.44;
Matches  109;    Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy    440 TTATAATTAATGATGATCTGAATAAACATACCAGGATAAAGATGTCAAATGAGTGTGACT 499
          ||||| |||||||||||||| ||||||||
Db   2309 TTAAATTAAGGAAAATTAGAAACAATCTTATTTTTTGAATATAGCAAATGCTACATTGC 2250

Qy    500 CCCTTAAAGTAGATTAAAGTGTGCATTCTTGTTTCCTAAAATATGATTTTACTGCTTGA 559
          ||||| ||||||
Db   2249 TAATTCAAGTATAAAACTTAAAAGATAATTTGTTTTATAAACACTTATTTGTTTACTTCT 2190

Qy    560 AATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTAAATAATGAATAATGCATGGAA 619
          ||||| ||||| ||||||
Db   2189 TAATTTTATTTAGTTATAATTAACTAAATAATAATAGCATTAAATATATTCTTTAAATTTA 2130

Qy    620 AATTATTATCCTTTGAAAACTGATTGATAAATATATTC 657
          ||||| |||||
Db   2129 AATTAATACTATATTATGACGGACACTTTCGTATTTTC 2092

RESULT 12
US-08-920-828-6/c
; Sequence 6, Application US/08920828
; Patent No. 5853998
; GENERAL INFORMATION:
;   APPLICANT: Ohno, Tsuneya
;   APPLICANT: Matsuhisa, Akio
;   APPLICANT: Uehara, Hirotsugu
;   APPLICANT: Eda, Soji
;   TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
```

```
RESULT 13
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;   APPLICANT: Dumas Milne Edwards, J.B.
;   APPLICANT: Jobert, S.
;   APPLICANT: Giordano, J.Y.
;   TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;   FILE REFERENCE: GENSET.054PR2
;   CURRENT APPLICATION NUMBER: US/09/621,976
;   CURRENT FILING DATE: 2000-07-21
```

```
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Best Local Similarity   51.0%;   Pred. No. 0.32;
Matches 102;   Conservative    0;   Mismatches    98;   Indels    0;   Gaps    0;

Qy    508 GTAGATTAAAGTGTGCATTCTTGTTCCTAAAATATGATTTACTGCTTGAAATTACAT 567
              |||   |||  ||||||   || ||||   ||||||  ||| ||||||||||||
Db   1508 GTCAATATATGGATATAGGCTTGATTTCGGCGATGCGTCGTTATCCCCGCTGAAATCACAC 1449

Qy    568 TTGAGTTGAAGTTAGAAACTAACATAGCATTAATAGAATAATGCATGGAAAATTATTA 627
              |||||||  ||   ||||  |||  ||||| ||||||||||  ||||||||||||
Db   1448 CGTGTTTTAGTCTATCACCAAAAGAAAGCAGTACTGGCAATCCAGCGGACACAAAAA 1389

Qy    628 TCCTTTGAAAACTGATTGATAAAATATATTCCCCTCCTTTAGAAACAGTCAAAAGGCACT 687
              |||||   || |||   |||||||||   ||||||||||||||||||||||||
Db   1388 GAGCCTCAAAATTGGTACATAAAATGTTTTGACCCTTTGGCTATTCAGTCAAAAGTCAGT 1329

Qy    688 TCAAACAAGTTCAAAATAA 707
              |||||||||||||||||||
Db   1328 CACGAACATTTACAACACAA 1309

RESULT 15
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match               2.6%;   Score 42.8;   DB 4;   Length 832;
Best Local Similarity   14.2%;   Pred. No. 0.33;
Matches  57;   Conservative 170;   Mismatches 172;   Indels   3;   Gaps    1;

Qy    929 TCCTATTATTCCCTTTGAAAACAAGAATAACAATAAGTTTAATAGTTGCCATACTTAG 988
              ||||||     ||   :  : :    :::::: :::  :    : :|:
Db    404 TGCTAAGTACAAGAGATTGTACCATAAATAATTATATTTTGTYTWTWWKTWWTYTWTYTWT 345

Qy    989 CATTTATCAGGTCTAATGAAACCAATATGAAATCTCGTGATAAAATATTTTCTGATGTTACT 1048
              || ::::||||::::  :  || | |:  ::  :   :: :::  :: :
Db    344 RMWWMKKGKARRWYTWWKSTYACASRYRKYTWGWWWYWYWKRMSTRWYCYMCWKCCMYRGR 285

Qy   1049 AGCTATGGGAAATTAGAACTGGCACAACCCTGACATTACTAAGTGGAAATGTTAGGATTT 1108
              ::  | |: ||   ::::::||::::|:  :| :|:   |    || |
Db    284 RCAWYTMARGRMWSYAWGKWKSMRSAMSMCTRMYYKKGSTYWTWMKCTCATWCYWYWKYWK 225

Qy   1109 TTCGGCATCGCATGT---TAGAATCTCTAAAATTTAAACATTCCTGTTAAATGACTAAGG 1165
              :   :: :|:  ||::  :|| | : |   ||   :::::   : :
Db    224 RMWSKTCWSGSRGGYMTSYTSTRSYSMYWASWMYTMCWWWGRWNSTYWMAWGKKWWRYA 165

Qy   1166 TTTGCTTTATCAATATGAATTCTGAAGGCCAATATCATCAATTAACTATGAAAGCTTT 1225
              ||::| :::|::: |:  :|:|  :  ::::    |: || :  : :
Db    164 TTWRRAMWMWAAWTMWYMWAWCMSSRGAAMYRRTMMWGYRYWWRRKSYRTRCAWA 105

Qy   1226 TAATTCCTAAAAAATAGTTTTAGAGATATTCAAGCAATGCTCTCCTAATATCCATACGCAA 1285
              :|:|:  | :: :   :::::| |  ::::|||::  ||:  :: | |
Db    104 YAWKTKRSYYWCWRNWKWKRCMWMMMMAMAYGKTMMRACWKTRYWRWAWAMWRMWWTMM 45

Qy   1286 GTGTGTTTATGACACAAATTCACTACGTCTGGTTTAAAAATGA 1327
              : :: | | |:|: :| :::|::::::
```

```
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match               2.7%;   Score 43.2;   DB 4;   Length 832;
Best Local Similarity   11.0%;   Pred. No. 0.27;
Matches  41;   Conservative 176;   Mismatches 154;   Indels   1;   Gaps    1;

Qy    551 ACTGCTTGAAATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTAATATGAATAA 610
              :::  |: |: ::::::::  :::::::   :::   :   :::: :|: : :
Db      3 WYWWKXTTWYAKCWTKWKNWSWSYMNYWKNYYMKTYWRWRRKKKKAWWKYWKTWTWYWRY 62

Qy    611 TGCATGGAAAATTATTATCCTTGAAAACTGATTGATAAAATATATTCCCCTCCTTTAGA 670
              :::|     :: :  :::::::::   :|:::  :::::::: :::::|||||||||||
Db     63 AMWGYYKKKAMCRTKYTKKKKKGYMWMWYWGWRRSYMAMWTRTWTGYAYYRSMYWWRYR 122

Qy    671 AACAGTCAAAAGCCACTTCAAACAAGTTTCAAAATAAAGGAAGTAGCAAGTTAGGCGAT 730
              ::  |: :  |: |:|: :::: :|::::: | ||:  ::::: :|:::::|::
Db    123 CWKKKAYYRKTTCYSSKGWTWWKRWKKAWTTWWWKCKTYYWAATRYWWMMCWTTKRWRASWW 182

Qy    731 GGATTATATTTTCTTGGCTTCTGTGTATACCCATTGGCCAGGGCCTTTATAAGGACTCCCA 790
              ::: |:   |||:|::|::| |:   :::: ::::::||:::::::
Db    183 YCWWWGKARKWSTWRKSRSYASARSAKRCCYSCSWGAMSWKYMWRMWRWGWATGAGMKA 242

Qy    791 AAAGCATTTGAAGAATG-GCAAATCAAATAAGTGTATGCTCCTCAAATGAGGCATTT 849
              ::|:|:  | |:||||| || ::::|:: :::::  :| ::|  :| |:|  :|
Db    243 WRASCMWRRKYAGKSKTSYKSMWMCWTRSWKYCYTKARWTGYYCYRKGGMWGKRGRWYAS 302

Qy    850 TTAATTGTTACAATCTTCATTTGGACGCTCAGGTTATGATATGTTTATGAAAAATAAGCTTC 909
              ::::::::|:::||:| |::  : :   :::   :::: | |  ||:: :
Db    303 KKYMWKRWWWCWARMYRYSTGTRASMWRWRYTYMMMKWWKYAWARAAWRWWAMWWAWRR 362

Qy    910 ATTATTCTTAT 921
              ||| || ||||
Db    363 ACAAAATATAAT 374

RESULT 14
US-09-202-893B-3/c
; Sequence 3, Application US/09202893B
; Patent No. 6319692
; GENERAL INFORMATION:
; APPLICANT: KADOTA, Mariko
; APPLICANT: KIWAKI, Mayumi
; APPLICANT: SAWAKI, Saeko
; APPLICANT: SHIRASAWA, Yukiko
; APPLICANT: SONE, Harue
; APPLICANT: SAKO, Tomoyuki
; TITLE OF INVENTION: METHODS FOR TRANSFERRING GENE INTO CHROMOSOME
; FILE REFERENCE: 980794/HG
; CURRENT APPLICATION NUMBER: US/09/202,893B
; CURRENT FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: PCT/JP97/02187
; PRIOR FILING DATE: 1997-06-25
; PRIOR APPLICATION NUMBER: JP 8/184266
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: JP 8/257764
; PRIOR FILING DATE: 1996-09-06
; NUMBER OF SEQ ID NOS: 9
SEQ ID NO 3
; LENGTH: 1542
; TYPE: DNA
; ORGANISM: Lactobacillus casei
US-09-202-893B-3

Query Match               2.7%;   Score 43.2;   DB 4;   Length 1542;
```

Mon Apr 26 09:48:37 2004

Db   44 MMYYNYWRAMKRRWWRKWRSWSWMWMAWGMTRWAARMWWRW 3

Search completed: April 26, 2004, 02:11:08
Job time : 88.3165 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:       April 24, 2004, 23:47:42 ; Search time 430.176 Seconds
                                          (without alignments)
                                          15988.422 Million cell updates/sec

Title:           US-10-051-835-15
Perfect score:   1619
Sequence:        1 ttaatccacatatcttatgt.........atttattaaaaatttaaaa 1619

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:    6747726

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  N_Geneseq_29Jan04:*
            1:  geneseqm1980s:*
            2:  geneseqm1990s:*
            3:  geneseqm2000s:*
            4:  geneseqm2001as:*
            5:  geneseqm2001bs:*
            6:  geneseqm2002s:*
            7:  geneseqm2003as:*
            8:  geneseqm2003bs:*
            9:  geneseqm2003cs:*
           10:  geneseqm2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1595	98.5	1619	8	ACH04252	Ach04252 Human CDN
2	477.8	29.5	618	7	ABX74725	Abx74725 Human CDN
c 3	475	29.3	580	6	ABL68031	Abl68031 Ovary can
c 4	475	29.3	580	6	ABL67459	Abl67459 Thyroid c
c 5	475	29.3	580	6	ABL67046	Abl67046 Thyroid c
c 6	475	29.3	580	6	ABL64323	Abl64323 Stomach c
7	52.6	3.2	6077	6	ABL33246	Abl33246 Human imm
8	51.8	3.2	8056	7	ABZ10100	Abz10100 Haematopo
9	51.2	3.2	8056	7	ABZ10246	Abz10246 Haematopo
c 10	50.6	3.1	9504	6	ABK28407	Abk28407 DNA trans
c 11	50.4	3.1	83391	6	ABQ67093	Abq67093 Human ang
c 12	49.8	3.1	32392	6	ABL56203	Abl56203 AmEPV gen
13	48.8	3.0	883	4	AAL15210	Aal15210 Human bre
c 14	48.8	3.0	8056	7	ABZ10246	Abz10246 Haematopo
c 15	48.6	3.0	6131	8	ABL32891	Abl32891 Human imm
c 16	48	3.0	8056	7	ABZ10100	Abz10100 Haematopo
c 17	47.2	2.9	1455	3	AAC50419	Aac50419 Arabidops
c 18	47.2	2.9	110000	4	AAV21209_03	Continuation (4 of
c 19	47	2.9	18585	6	ABL34609	Abl34609 Human met
20	46.8	2.9	2000	7	ADA71938	Ada71938 Rice gene
21	46.8	2.9	6641	6	ABL32315	Abl32315 Human imm
22	46.8	2.9	6641	6	ABL54336	Abl54336 Chemical1
23	46.8	2.9	10957	6	ABL33111	Abl33111 Human imm
c 24	46.6	2.9	6247	6	ABK39923	Human che
c 25	46.6	2.9	8170	4	ABK28258	DNA trans
26	46.6	2.9	13469	4	AAS46594	Tumour su
27	46.6	2.9	15479	6	ABK39964	Human che
c 28	46.2	2.9	5771	6	ABL33951	Human imm
c 29	46.2	2.9	15373	6	ABL32466	Human imm
30	45.6	2.8	5499	6	ABQ66971	Human ang
31	45.6	2.8	7456	6	ABL33931	Human imm
32	45.6	2.8	7456	6	ABL92293	Chemical1
c 33	45.6	2.8	8771	4	ABL33824	Human imm
34	45.6	2.8	9706	6	AAK86270	Human imm
35	45.4	2.8	15479	6	ABK39965	Human che
36	45.4	2.8	4704	2	AAQ05264	Sequence
37	45.4	2.8	11049	6	ABL32669	Human imm
38	45.4	2.8	11049	6	ABL92219	Chemical1
39	45.4	2.8	11049	6	ABL49322	Human pol
40	45.4	2.8	17594	6	ABL34026	Human imm
41	45.2	2.8	11787	9	ADB54330	Pretreate
42	45.2	2.8	15825	6	ABL33242	Human imm
43	45.2	2.8	15825	6	ABL34558	Human met
44	45	45	6888	3	AAA70114	Plasmodiu
45	45	45	18154	6	ABL32254	Human imm

ALIGNMENTS

```
RESULT 1
ACH04252
ID   ACH04252 standard; cDNA; 1619 BP.
XX
AC   ACH04252;
XX
DT   26-SEP-2003  (first entry)
XX
DE   Human cDNA differentially expressed in lung cancer #457.
XX
KW   Gene therapy; emphysema; ss; gene; chronic obstructive pulmonary disease;
KW   respiratory disorder; lung cancer; asthma; human.
XX
OS   Homo sapiens.
XX
PN   US2003065157-A1.
XX
PD   03-APR-2003.
XX
PF   04-APR-2002; 2002US-00116802.
PR   04-APR-2001; 2001US-0281593P.
XX
PA   (LASE/) LASEK A W.
XX
PI   Lasek AW;
XX
DR   WPI; 2003-540803/51.
XX
PT   New combination comprising cDNAs that are differentially expressed in
PT   respiratory disorders, useful for diagnosing or treating respiratory
PT   disorders e.g., lung cancer, chronic obstructive pulmonary disease,
PT   emphysema or asthma.
XX
PS   Claim 1; Page; 39pp; English.
XX
CC   The invention relates to a combination comprising cDNAs or their
CC   complements that are differentially expressed in respiratory disorder.
CC   The combination is useful for preparing a composition for diagnosing or
CC   treating respiratory disorders e.g. lung cancer, chronic obstructive
CC   pulmonary disease, emphysema or asthma. The present sequence represents
CC   human cDNA differentially expressed during lung cancer
XX
SQ   Sequence 1619 BP; 560 A; 240 C; 287 G; 508 T; 0 U; 24 Other;

Query Match             98.5%;  Score 1595;  DB 8;  Length 1619;
```

Best Local Similarity 100.0%; Pred.No. 0;
Matches 1619; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy       1 TTAATCACTATATCCTATGTATAATATTGCCTAATCCATAGAAAAAGGATATAAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTAATCACTATATCCTATGTATAATATTGCCTAATCCATAGAAAAAGGATATAAAG 60

Qy      61 TATTAAATATGTGATATATAGCNNNNNNNNNNNNNNNNNATAGGAAGTTCAA 120
           ||||||||||||||||||||||              |||||||||||||||||
Db      61 TATTAAATATGTGATATATAGCNNNNNNNNNNNNNNNNNATAGGAAGTTCAA 120

Qy     121 GTCACTCAATTGAAGAAACATATCTCTGAGCATAGGAGCAGCCTCCAGGTCCTATGGTGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTCACTCAATTGAAGAAACATATCTCTGAGCATAGGAGCAGCCTCCAGGTCCTATGGTGG 180

Qy     181 GATGCAGTGGACAGCAGGAGGGGGAAATTAGAAAAGAGAACTATATAATTGAAAAAGGAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GATGCAGTGGACAGCAGGAGGGGGAAATTAGAAAAGAGAACTATATAATTGAAAAAGGAT 240

Qy     241 ATAAAGCATTAAATATATGATATATAGCTATATCTATGTATGTATCTAACAGAGAAGTTC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ATAAAGCATTAAATATATGATATATAGCTATATCTATGTATGTATCTAACAGAGAAGTTC 300

Qy     301 AGTCACTTCAATTAAAGAAACAATTTTGAGCATGGACCAGCCTCAGGTCTTATGCTGG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGTCACTTCAATTAAAGAAACAATTTTGAGCATGGACCAGCCTCAGGTCTTATGCTGG 360

Qy     361 GATGCAGTAGACAGGAGATGGGAAATTAGAAAAGAGAACTGTGTAATTGAAATGACGTG 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GATGCAGTAGACAGGAGATGGGAAATTAGAAAAGAGAACTGTGTAATTGAAATGACGTG 420

Qy     421 GGCTGCACCCTTAAGAACTTATAATAATGATCTGAATAAACATACCAGGATAAAG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGCTGCACCCTTAAGAACTTATAATAATGATCTGAATAAACATACCAGGATAAAG 480

Qy     481 ATGTCAAATGAGTGTGACTCCCTAAAGTAGAGTAAAGTGTGCATTCTTTGTTCCTAAA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATGTCAAATGAGTGTGACTCCCTAAAGTAGAGTAAAGTGTGCATTCTTTGTTCCTAAA 540

Qy     541 ATATGATTTTACTGCTTGAAATTACATTTGAAGTTGAAGTTAGAAACATAGCATTA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATATGATTTTACTGCTTGAAATTACATTTGAAGTTGAAGTTAGAAACATAGCATTA 600

Qy     601 ATATGAATAATGCATGGAAATTATTATCCTTTGAAAACTGATAATAAATATATTCCCC 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATATGAATAATGCATGGAAATTATTATCCTTTGAAAACTGATAATAAATATATTCCCC 660

Qy     661 CTCCCTTAGAAACAGTCAAAAGCCACTTCAAACAAGTTTCAAAATAAAGGAAGGTAGCAA 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CTCCCTTAGAAACAGTCAAAAGCCACTTCAAACAAGTTTCAAAATAAAGGAAGGTAGCAA 720

Qy     721 GTTAGGCGATGGATTATATTTCTTGGCTTGTGTATACCCATTGGCCAGGGGCCTTTATA 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GTTAGGCGATGGATTATATTTCTTGGCTTGTGTATACCCATTGGCCAGGGGCCTTTATA 780

Qy     781 AGGACTCCCAAAAGCATTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCTCAAAT 840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 AGGACTCCCAAAAGCATTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCTCAAAT 840

Qy     841 GAGGGCATTTTAATTGTTACAATCTATTTGGACGCGCTCAGGTTATGATATGTTTATGAAAA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GAGGGCATTTTAATTGTTACAATCTATTTGGACGCGCTCAGGTTATGATATGTTTATGAAAA 900

Qy     901 AATAAGTTCATTATTTCTTATAGCTACATCCTATTATTCCCTTTAGAAACAAGAATAAC 960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AATAAGTTCATTATTTCTTATAGCTACATCCTATTATTCCCTTTAGAAACAAGAATAAC 960

Qy     961 AATAAGTTTTAATAGTTGCCATACTTAGCATTTATCAGGTTCTAATGAAAACCAATATTGAA 1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AATAAGTTTTAATAGTTGCCATACTTAGCATTTATCAGGTTCTAATGAAAACCAATATTGAA 1020

Qy    1021 TCTCTGATAAATATTTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TCTCTGATAAATATTTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTG 1080

Qy    1081 ACATTACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCTAAAATTT 1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ACATTACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCTAAAATTT 1140

Qy    1141 AAACATTCCTGTTAAATGACTAAGGTTGCTTTATCAATGAATTCTGAAGGCCAATA 1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 AAACATTCCTGTTAAATGACTAAGGTTGCTTTATCAATGAATTCTGAAGGCCAATA 1200

Qy    1201 TCATACCATTAACTATGAAAGCTTTAATTCCTAAAAAATAGTTTTAGAGATATTCAAGCA 1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TCATACCATTAACTATGAAAGCTTTAATTCCTAAAAAATAGTTTTAGAGATATTCAAGCA 1260

Qy    1261 ATGCTTCTCCTAATATCCATACGCAAGTGTGTTTATGACACAAATTCACTAGTCTCGGTTTA 1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ATGCTTCTCCTAATATCCATACGCAAGTGTGTTTATGACACAAATTCACTAGTCTCGGTTTA 1320

Qy    1321 AAAATGAAATCTTATATTGACTGGGTGTCCCACATATTCCAGTAATTCTGTTATGAGA 1380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 AAAATGAAATCTTATATTGACTGGGTGTCCCACATATTCCAGTAATTCTGTTATGAGA 1380

Qy    1381 GGACTTGAAATAGCAAATTGCCCACAGTTAACTGATAGATCAGTGGTGGTGATCA 1440
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GGACTTGAAATAGCAAATTGCCCACAGTTAACTGATAGATCAGTGGTGGTGATCA 1440

Qy    1441 TAACCACTTGGTACTACACCCCAGAAAACTCAAAATTGTCTTTTCTCCCTGATGAGATATGGG 1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 TAACCACTTGGTACTACACCCCAGAAAACTCAAAATTGTCTTTTCTCCCTGATGAGATATGGG 1500

Qy    1501 GTGTCCTTTGTACGTCAGGGCTCAGGGCTCAGGTACCCAAGTGAAGTGAATATAAGCAAAT 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 GTGTCCTTTGTACGTCAGGGCTCAGGGCTCAGGTACCCAAGTGAAGTGAATATAAGCAAAT 1560

Qy    1561 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAATTTAAAA 1619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAATTTAAAA 1619
```

RESULT 2
ABX74725/c
ID   ABX74725 standard; cDNA; 618 BP.
XX
AC   ABX74725;
XX
DT   21-MAR-2003 (first entry)
XX
DE   Human cDNA sequence #135 down-regulated in CC-RCC patients.
XX
KW   Human; microarray; solid surface; immobilised probe; CC-RCC;
KW   differential expression profile; aggressive CC-RCC tumour type;
KW   non-aggressive CC-RCC tumour type; clear cell renal carcinoma;
KW   gene expression profiling; tumour tissue; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200279411-A2.
XX
PD   10-OCT-2002.
XX
PF   29-MAR-2002; 2002WO-US009576.
XX
PR   29-MAR-2001; 2001US-0279411P.
XX
PA   (VAND-) VAN ANDEL INST.
XX
PI   Haab B, Rhodes D, Teh BT, Takashi M;
XX
DR   WPI; 2003-040679/03.
XX
PT   New microarray, comprising a matrix of cDNA probe from a set of probes
PT   immobilized to a solid surface in predetermined order, useful in the
PT   prognosis of patients with clear cell renal carcinoma.

XX PS Example 3; SEQ ID NO 295; 179pp; English.
XX
CC The present invention relates to a microarray comprising a matrix of at
CC least one cDNA probe from a set of probes immobilised to a solid surface
CC in a predetermined order, where a row of pixels corresponds to replicates
CC of one distinct probe from the set. The probes are complementary to
CC nucleic acid sequences that are expressed differentially in aggressive as
CC compared to non-aggressive types of clear cell renal carcinoma (CC-RCC)
CC and that hybridise to the probes under high stringency conditions. The
CC microarray is useful for the prognosis of patients with CC-RCC, wherein
CC aggressive and non-aggressive CC-RCC tumour types are characterised by
CC differential expression profiles of genes that hybridise with one or more
CC probes immobilised on the microarray. The arrays are useful for gene
CC expression profiling of tumour and normal tissues. The present sequence
CC represents a human cDNA sequence down-regulated in CC-RCC patients
XX
SQ Sequence 618 BP; 203 A; 106 C; 106 G; 197 T; 0 U; 6 Other;

Query Match   29.5%;   Score 477.8; DB 7; Length 618;
Best Local Similarity 93.6%;   Pred. No. 1.2e-95;
Matches 573; Conservative   0; Mismatches   30; Indels   9; Gaps   7;

```
Qy   1008 AACCAATATTGAATCTCTGATAAATATTTTCTGATGTTACTAGCTATGGGAAATTAGAAC 1067
               ||||| || |||||||||  ||| ||||||| || ||||||||||||||||||||||
Db    603 AACCCATATGGAACNCGGATAAATATTTTCCGGANGTACTNGCTATGGGAAATTGGAAC  544

Qy   1068 TGGCACAACCCTGACATTACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGA 1127
               |||||| ||| ||||||||||||||||||||| ||||| ||||||||||||||||||
Db    543 TGGCACACCCCTGACATTACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGA  484

Qy   1128 ATCTCTAAAATTTAAACATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATATGAATT 1187
               |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    483 ATCTCTAAAATTTAAACATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATATGAATT  424

Qy   1188 CTGAAGGCCAATATCATACCAATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTTAG 1247
               ||||||||| ||||||||||||||||||||||||||| |||||||||||||||||||||
Db    423 CTGAAGGCCAATATCATACCAATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTTAG  364

Qy   1248 AGATATTCAAGCAATGCTCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCA 1307
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 AGATATTCAAGCAATGCTCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCA  304

Qy   1308 CTAGTCTGGTTAAAAATGAAATCTTTATATTGACTAGCGGGGTCCCACATATTCCAGTAAT 1367
               ||||||| |||||||||||||||||||||||  ||  ||||||||||||||||||||||||
Db    303 CTAGTCT-GTTAAAAATGAAATTCTTTATATTGACT-GGTGTTCCACATATTCCAGTAAT  246

Qy   1368 TTCTGTTTATGAGAGGACTTGAAATAGCAAATTGCCCACACAGTTAACTGGATAGATCACG 1427
               ||||||||||||||||||||||||||||||||  |||||| ||||||||||||| |||||
Db    245 TTCTGTTTATGAGAGGACTTGAAATAGCAAATTG-CCACACAGTTAACTGGATATGACACG  187

Qy   1428 TACGTGGTGGATCAATAACCACTTGGTACTACACCCAGAAACTCAAAATGTCTTTTCTCCT 1487
               ||||||||||||||||||||||||||||||||||||||||||||||  || |||||||||
Db    186 TACGTGGTGGATCAATAACCACTTGGTACTACACCCAGAAACTCAAAATTGTC-TTTCTCCT  128

Qy   1488 GATGAGATATGGGGTGTCCTTTGTACGTCTAGGGCCTAGGGTACCCAAGTGAAGTGAAT 1547
               ||||||||||| |||||||||||||||||||||||          |||  ||||||||
Db    127 GATGAGATAT-GGGTGTCCTTTGTACGTCTAGGGCC---TAGGTAACCAGTGGAGTGATT   72

Qy   1548 ATATAAGCAAAATGTGTTTGTATCCGAGTCTTCCTGTCATTGTAATAAAAAAATTTATTT 1607
               ||||||| ||||||||||||||||||||||||||||||||||||||| |||||||||||
Db     71 ATATTAGC-AAATGTGTTTGTATCCGAGTCTTCCTGTCATTGTAATAAAAAATTTATTT   13

Qy   1608 AAAAATTTAAA 1619
               ||||||||||
Db     12 AAAAATTTAAA    1
```

RESULT 3
ABL68031/c
ID ABL68031 standard; DNA; 580 BP.
XX
AC ABL68031;
XX
DT 15-MAY-2002 (first entry)
XX
DE Ovary cancer related gene sequence SEQ ID NO:6368.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0231133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235114P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 02-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237599P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
PS Claim 1; SEQ ID NO 6368; 44pp; English.

```
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 580 BP; 198 A; 98 C; 96 G; 187 T; 0 U; 1 Other;

Query Match             29.3%;   Score 475;   DB 6;   Length 580;
Best Local Similarity   95.7%;   Pred. No. 5e-95;
Matches   563;   Conservative   0;   Mismatches   16;   Indels   9;   Gaps   7;

Qy   1025 TGATAAATATTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACAT 1084
          ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    580 TGATANATATTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACAT 521

Qy   1085 TACTAAGTGGAAAATGTTAGGATTTTTCGGCATCGCATGTTAGAATCTCAAAATTTAAAC 1144
          ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    520 TACTAAGTGGAAAATGTTAGGATTTTTCGGCATCGCATGTTAGAATCTCAAAATTTAAAC 461

Qy   1145 ATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCAT 1204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    460 ATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCAT 401

Qy   1205 ACCATTAACTATGAAAGCTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCAATGC 1264
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    400 ACCATTAACTATGAAAGCTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCAATGC 341

Qy   1265 TCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCTCGGTTTAAAAA 1324
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    340 TCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTC-GTTTAAAAA 282

Qy   1325 TGAAATCTTTATATTGACTGGGTGTCCCACATATTCCAGTAATTCTGTTATGAGGAGGAC 1384
          |||||||||||||||||||| |||||||||||||||||||||||||| ||||||||||||
Db    281 TGAATTCTTTATATTGACT-GGTGTTCCCACATATTCCAGTAATTTCTGTTATGAGGAGGAC 223

Qy   1385 TTGAAATAGCAAATTGCCCACACAGTTAACTGGATAGATCACGTACGTGGTGATCATAAC 1444
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    222 TTGAAATAGCAAATTG-CCACACAGTTAACTGGATAGATCGATAGACCACGTGGTGATCATAAC 164

Qy   1445 CACTTGGTACTACACCCAGGAAACTCAAAATTGTCTTTCTCCTGATGAGATATGGGGTGT 1504
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||| |||
Db    163 CACTTGGTACTACACCCAGGAAACTCAAAATTGTC-TTTCTCTGATGAGATAT-GGGTGT 106

Qy   1505 CCTTTTGTACGTCTAGGGCCTAGGGTACCCAAGGTACCCAAGTGAAGTGAATATATAAGCAAAATGTGT 1564
          ||||||||||||||||||||||||   |||||||||||||||||||||| ||||||||||
Db    105 CCTTTTGTACGTCTAGGGCCC---TAGGTAACCAGTGGAGTGAATATTAGC-AAATGTGT 50

Qy   1565 TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAA 1612
          ||||||||||||||||||||||||||||||| |||||||||||||||
Db     49 TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAA 2

RESULT 4
ABL67459/c
ID   ABL67459 standard; DNA; 580 BP.
XX
AC   ABL67459;
XX
DT   15-MAY-2002  (first entry)
```

XX
DE   Thyroid cancer related gene sequence SEQ ID NO:5796.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO20019462 9-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 5796; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-

```
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 580 BP; 198 A; 98 C; 96 G; 187 T; 0 U; 1 Other;

Query Match             29.3%;  Score 475;  DB 6;  Length 580;
Best Local Similarity   95.7%;  Pred. No. 5e-95;
Matches   563;  Conservative    0;  Mismatches   16;  Indels    9;  Gaps    7;

Qy     1025 TGATAAATATTTCTGATGTTACTAGCTANGGGAAATTAGAACTGGCACAACCCTGACAT 1084
            ||||| |||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      580 TGATANATATTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACAT 521

Qy     1085 TACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGATATCTTAAAATTTAAAC 1144
            |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db      520 TACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGATATCTTAAAATTTAAAC 461

Qy     1145 ATTCCTGTAAATGACTAAGGTTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCAT 1204
            |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db      460 ATTCCTGTAAATGACTAAGGTTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCAT 401

Qy     1205 ACCATTAACTATGAAAGCTTTAATTCCTAAAAAATAGTTTTAGAGATATTCAAGCAATGC 1264
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      400 ACCATTAACTATGAAAGCTTTAATTCCTAAAAAATAGTTTTTAGAGATATTCAAGCAATGC 341

Qy     1265 TCTCCTAAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCTGGTTTAAAAA 1324
            |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db      340 TCTCCTAAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCT-GTTTAAAAA 282

Qy     1325 TGAAATCTTATATTGACTGGGTGTCCCACATATTCCAGTAATTCTCGTTATGAGAGGAC 1384
            |||| ||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db      281 TGAATTCTTTATATTGACT-GGTGTTCCACATATTTCAGTAATTCTGTTATGAGAGGAC 223

Qy     1385 TTGAAATAGCAAATGCCCACACAGTTAACTGGATAGATCCGATAGATCACGTGGTCATAAC 1444
            |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      222 TTGAAATAGCAAATG-CCACACAGTTAACTGGATAGACCACGTGGTGTCATCATAAC 164

Qy     1445 CACTTGGTACTACACCCCAGAAACTCAAAATTGTCTTTCTCCTGATGAGATATGGGGTGT 1504
            |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      163 CACTTGGTACTACACCCCAGAAACTCAAAATTGTC-TTCTCCTGATGAGATAT-GGGTGT 106

Qy     1505 CCTTTTGTACGTCTAGGGCCTAGGGCCCAAGTGACCCAAGTGAATATATAAGCAAAATGTGT 1564
            |||||||||||||||||||||||||    |||||||||||||||||||||||||||||||||
Db      105 CCTTTTGTACGTCTAGGGCCTAGGCC---TAGTAACCAGTGGAGTGATTATAGCC-AAATGTGT 50

Qy     1565 TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAATTTATTTAAAAA 1612
            ||| |||||||||||||||||| |||||||||| |||||||||||||
Db       49 TTGTATCCAGAGTCTTCCTGCATTGTAATAAAAAAATTTATTTAAAAA 2
```

RESULT 5
ABL67046/c
ID   ABL67046 standard; DNA; 580 BP.
XX
AC   ABL67046;
XX
DT   15-MAY-2002 (first entry)
XX
DE   Thyroid cancer related gene sequence SEQ ID NO:5383.

XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US010838.
XX
PR   05-JUN-2000; 2000US-0209473P.
PR   05-JUN-2000; 2000US-0209531P.
PR   18-SEP-2000; 2000US-0233133P.
PR   18-SEP-2000; 2000US-0233617P.
PR   20-SEP-2000; 2000US-0234009P.
PR   20-SEP-2000; 2000US-0234034P.
PR   20-SEP-2000; 2000US-0234052P.
PR   22-SEP-2000; 2000US-0234509P.
PR   22-SEP-2000; 2000US-0234567P.
PR   25-SEP-2000; 2000US-0234923P.
PR   25-SEP-2000; 2000US-0234924P.
PR   25-SEP-2000; 2000US-0235077P.
PR   25-SEP-2000; 2000US-0235082P.
PR   25-SEP-2000; 2000US-0235134P.
PR   25-SEP-2000; 2000US-0235280P.
PR   26-SEP-2000; 2000US-0235637P.
PR   26-SEP-2000; 2000US-0235638P.
PR   27-SEP-2000; 2000US-0235711P.
PR   27-SEP-2000; 2000US-0235720P.
PR   27-SEP-2000; 2000US-0235840P.
PR   27-SEP-2000; 2000US-0235863P.
PR   28-SEP-2000; 2000US-0236028P.
PR   28-SEP-2000; 2000US-0236032P.
PR   28-SEP-2000; 2000US-0236033P.
PR   28-SEP-2000; 2000US-0236034P.
PR   28-SEP-2000; 2000US-0236109P.
PR   28-SEP-2000; 2000US-0236111P.
PR   29-SEP-2000; 2000US-0236842P.
PR   29-SEP-2000; 2000US-0236891P.
PR   02-OCT-2000; 2000US-0237172P.
PR   02-OCT-2000; 2000US-0237173P.
PR   02-OCT-2000; 2000US-0237278P.
PR   02-OCT-2000; 2000US-0237294P.
PR   02-OCT-2000; 2000US-0237295P.
PR   02-OCT-2000; 2000US-0237316P.
PR   03-OCT-2000; 2000US-0237425P.
PR   03-OCT-2000; 2000US-0237598P.
PR   03-OCT-2000; 2000US-0237604P.
PR   03-OCT-2000; 2000US-0237606P.
PR   03-OCT-2000; 2000US-0237608P.
PR   01-NOV-2000; 2000US-0244867P.
PR   01-NOV-2000; 2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 5383; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in

```
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. MI can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of MI, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. MI can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 580 BP; 198 A; 98 C; 96 G; 187 T; 0 U; 1 Other;

Query Match             29.3%;  Score 475;  DB 6;  Length 580;
Best Local Similarity   95.7%;  Pred. No. 5e-95;
Matches  563;  Conservative   0;  Mismatches  16;  Indels   9;  Gaps    7;

Qy   1025 TGATAAAATATTTTCTGATGTTACTAGCTATGGAAATTAGAACTGGCACAACCCTGACAT 1084
               |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    580 TGATANATATTTTCTGATGTTACTAGCTATGGAAATTAGAACTGGCACAACCCTGACAT  521

Qy   1085 TACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCTAAAATTTAAAC 1144
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    520 TACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCTAAAATTTAAAC  461

Qy   1145 ATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCAT 1204
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    460 ATTCCCGTTAAATGACTAAGGTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCAT  401

Qy   1205 ACCATTAACTATGAAAGCTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCAATGC 1264
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    400 ACCATTAACTATGAAAGCTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCAATGC  341

Qy   1265 TCTCCTAAATATCCATACGCAAGTGTGTTTATGACACAAATTCACTAGTCTGGTTAAAA 1324
               ||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    340 TCTCCTAAATATCCATACGCAAGTGTGTTTATGACACAAATTCACTAGTCT-GTTAAAA  282

Qy   1325 TGAAATCTTTATATTGACTGGGTGTCCCACATATTCCAGTAATTTCTGTTATGAGAGGAC 1384
               |||||||||||||||||| || ||| ||||||||||||||||||||||||||||
Db    281 TGAAATCTTTATATTGACT-GGTGTTCCACATATTCCAGTAATTTCTGTTATGAGAGGAC  223

Qy   1385 TTGAAATAGCAAATTGCCCACCAGAGTAACTGGATAGATCACGTGGTGGTCGTGATCATAC 1444
               ||||||||||||||| |||||| |||||||||||||||||| | |||| ||||||||
Db    222 TTGAAATAGCAAATTG-CCACCAGTTAACTGGATAGATCACCACGTAGTCGTGATCATAAC  164

Qy   1445 CACTTGGTACTACACCCCAGAAACTCAAAATTGTCTTTCTCCTGATGAGATATGGGGTGT 1504
               |||||||||||||| |||||||||||||||| || |||||||||||||||| ||||||
Db    163 CACTTGGTACTACACACCCCAGAAACTCAAAAATGTC-TTTCTCCTGATGAGATAT-GGGTGT  106

Qy   1505 CCTTTTGTACGTCTAGGGCCTCAGGGTACCCAAGTGAAGTGAATATAAGCAAAATGTGT 1564
               |||||||||||||||| ||   ||||| |||| ||| ||||| ||| || |||||||||
Db    105 CCTTTTGTACGTCTAGGCC---TAGGTAACCAGTGGAGTGATTATTAGC-AAATGTGT   50

Qy   1565 TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAA 1612
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db     49 TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAA    2
```

```
RESULT 6
ABL64323/c
ID   ABL64323 standard; DNA; 580 BP.
XX
AC   ABL64323;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Stomach cancer related gene sequence SEQ ID NO:2660.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO20019629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001;  2001WO-US010838.
XX
PR   05-JUN-2000;  2000US-0209473P.
PR   05-JUN-2000;  2000US-0209531P.
PR   18-SEP-2000;  2000US-0233113P.
PR   18-SEP-2000;  2000US-0233617P.
PR   20-SEP-2000;  2000US-0234009P.
PR   20-SEP-2000;  2000US-0234034P.
PR   20-SEP-2000;  2000US-0234052P.
PR   22-SEP-2000;  2000US-0234509P.
PR   22-SEP-2000;  2000US-0234567P.
PR   25-SEP-2000;  2000US-0234923P.
PR   25-SEP-2000;  2000US-0234924P.
PR   25-SEP-2000;  2000US-0235077P.
PR   25-SEP-2000;  2000US-0235082P.
PR   25-SEP-2000;  2000US-0235134P.
PR   25-SEP-2000;  2000US-0235280P.
PR   26-SEP-2000;  2000US-0235637P.
PR   26-SEP-2000;  2000US-0235638P.
PR   27-SEP-2000;  2000US-0235711P.
PR   27-SEP-2000;  2000US-0235720P.
PR   27-SEP-2000;  2000US-0235840P.
PR   27-SEP-2000;  2000US-0235863P.
PR   28-SEP-2000;  2000US-0236028P.
PR   28-SEP-2000;  2000US-0236032P.
PR   28-SEP-2000;  2000US-0236033P.
PR   28-SEP-2000;  2000US-0236034P.
PR   28-SEP-2000;  2000US-0236109P.
PR   28-SEP-2000;  2000US-0236111P.
PR   29-SEP-2000;  2000US-0236842P.
PR   29-SEP-2000;  2000US-0236891P.
PR   02-OCT-2000;  2000US-0237172P.
PR   02-OCT-2000;  2000US-0237173P.
PR   02-OCT-2000;  2000US-0237278P.
PR   02-OCT-2000;  2000US-0237294P.
PR   02-OCT-2000;  2000US-0237295P.
PR   02-OCT-2000;  2000US-0237316P.
PR   02-OCT-2000;  2000US-0237425P.
PR   03-OCT-2000;  2000US-0237598P.
PR   03-OCT-2000;  2000US-0237604P.
PR   03-OCT-2000;  2000US-0237606P.
PR   03-OCT-2000;  2000US-0237608P.
PR   01-NOV-2000;  2000US-0244867P.
PR   01-NOV-2000;  2000US-0245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a chemical
PT   agent to be tested for anti-neoplastic activity, and determining a change
PT   in expression of a gene of a signature gene set.
XX
PS   Claim 1; SEQ ID NO 2660; 44pp; English.
XX
CC   The present invention describes a method (MI) for screening for an anti-
CC   neoplastic agent. The method involves exposing cells to a chemical agent
CC   to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
```

```
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening an
CC   anti-neoplastic agent, and can be used for producing a product which is
CC   the data collected with respect to the anti-neoplastic agent as a result
CC   of M1, and the data is sufficient to convey the chemical structure and/or
CC   properties of the agent. M1 can be used in the treatment of cancer such
CC   as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC   prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC   cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC   cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC   tumour
XX
SQ   Sequence 580 BP; 198 A; 98 C; 96 G; 187 T; 0 U; 1 Other;

Query Match             29.3%;  Score 475;  DB 6;  Length 580;
Best Local Similarity   95.7%;  Pred. No. 5e-95;
Matches  563;           Conservative   0;  Mismatches  16;  Indels   9;  Gaps   7;

Qy  1025  TGATAAATATTTCTGATGTTACTAGCTATGGAAATTAGAACTGGCACAACCCTGACAT 1084
          |||| ||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   580  TGATANATATTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACAT 521

Qy  1085  TACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTGTTAGAATCTCTAAAATTTAAAC 1144
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   520  TACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTGTTAGAATCTCTAAAATTTAAAC 461

Qy  1145  ATTCCTGTTAAATGACTAAGTTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCAT 1204
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   460  ATTCCTGTTAAATGACTAAGTTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCAT 401

Qy  1205  ACCATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTTAGAGATATTTCAAGCAATGC 1264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   400  ACCATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTTAGAGATATTTCAAGCAATGC 341

Qy  1265  TCTCCAATATCCATACGCAAGTGTGTTTATGACACAAATTCACTAGTCTGGTTTAAAAA 1324
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   340  TCTCCAATATCCATACGCAAGTGTGTTTATGACACAAATTCACTAGTCT-GTTTAAAAA 282

Qy  1325  TGAAATCTTTATATTCACTGGGTGTCCCACATATTCCAGTAATTTCTGTTATGAGAGGAC 1384
          |||| |||||||||| ||| ||||| |||||||| ||||||||||||||||||||||||
Db   281  TGAATTCTTTATATTGACT-GGTGTTCCACATATTTCAGTAATTTCTGTTATGAGAGGAC 223

Qy  1385  TTGAAATAGCAAAATTGCCCCACACAGTTAACTGGATAGATCACGTACGTGGTGATCATAAC 1444
          ||||||||||||||||| |||||||||||||||||||| |||||||||||||||||||||
Db   222  TTGAAATAGCAAAATTG-CCACACAGTTAACTGGATAGACCACGTACGTGGTGATCATAAC 164

Qy  1445  CACTTCGTACTACACCCAGAAACTCAAAAATTGTCTTTTCTCCTGATGAGATATGGGGTGT 1504
          |||||||||||||||||||||||||||||||||| |||||||||||||||||| ||||||
Db   163  CACTTCGTACTACACCCAGAAACTCAAAAATTGTC-TTTCTCCTGATGAGATAT-GGGTGT 106

Qy  1505  CCTTTTGTACGTCTAGGGCCTAGGGTACCCAAGTGAAGTGAATATATAAGCAAAATGTGT 1564
          |||||||||||||||||           ||||| ||||| ||| |||| ||||||||||
Db   105  CCTTTTGTACGTCTAGGCC---TAGGTAACCAGTGGAGTGATTATTAGC-AAATGTGT 50

Qy  1565  TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAATTTATTTAAAAA 1612
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    49  TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAATTTATTTAAAAA 2
```

RESULT 7
ABL33246
ID   ABL33246 standard; DNA; 6077 BP.
XX
AC   ABL33246;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1219.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01033529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1219; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6077 BP; 1524 A; 185 C; 1391 G; 2977 T; 0 U; 0 Other;

```
Query Match             3.2%;   Score 52.6;  DB 6;  Length 6077;
Best Local Similarity   44.9%;  Pred. No. 0.098;
Matches  199;           Conservative   0;  Mismatches  244;  Indels   0;  Gaps   0;

Qy   809  GCAATATCAAATAAGTGTATGTCCCTCCAAATGAGGCATTTTAATTGTTACAATCATT 868
          ||  ||| | ||||| ||  || |||  ||| ||||| ||  ||||  | | ||| |||
Db    39  GGATTATTTAGAAAGTGATTGAGTTAGTGTTTAATTTTAATTTTATATTTTTGT 98

Qy   869  TGGACGCTCAGGTTATGATATGTTTATGAAAAATAAGCTTCATTATTCTTATAGCTACA 928
          |||  |  | |   |  ||||| || | ||  |||| | |    |||| ||| ||| ||
Db    99  GAAAAAGTTTATATTTGTATAAATTAATTAATTTAATTGGGTTTTTGTTTTGTTGGGAGTTAAA 158

Qy   929  TCCTATTATTCCCTTTAGAAACAAGAATAACAATAAGTTTAATAGTTGCCATACTTAG 988
          | ||||| ||||  |||||   | | ||| ||| | || ||||| || | | |||||
Db   159  AATTATTTTTAATAGTTATTGTTAAAAATGTATAAAAATTTGTTTGTTCGTAATATTTTT 218

Qy   989  CATTTATCAGGTCAATGAAACCAATATGAATCTCTGATAAATATTTCTGATGTTACT 1048
          | |||||| ||| |||| | | ||||||||| | | || |||| ||||| ||| ||
Db   219  ATTTTATATTTTTTTTTTTAGAATAGTTTTATTAATTTTATTAGGTTTAATTGAGGATGTT 278

Qy  1049  AGCTATGGGAAAATTAGAACTGGCACAACCCTGACATTACTAAGTCTAAGTGGAAATGTTAGGATTT 1108
          |  ||| |  ||||| ||| | |||| | |||| ||||| |||   |||| | ||| ||||| |||
Db   279  AAGTATAAATTATTAGTAGTTAGAATATAGTAATAAAATTATAAACATTCCTGTTAAAATGACTAAGTT 338

Qy  1109  TTCGGCATCGCATGTTAGAATCTCAAAATTAAAAATTAATTAAAATGACTAAGGTTT 1168
          || | |  |  |||| | || || |    ||    ||||| |  |   | |   | ||
Db   339  TTTGATTTGTGTGTTTGATTTATTATTATTATTAGATTTATATTATTTATTTTTAAAAATT 398

Qy  1169  GCTTTTATCAATATGAATTCTGAAGGCCAATATCATCCATTACTATGAAAGCTTTTAA 1228
          | ||||| | |||||||| | |   | | |||| ||| | ||| || || | ||||||
Db   399  TTATATTTGAAAAATAATTATTATATATTATTATTATTTTTGGAGATTTATTGGATGTATATAA 458

Qy  1229  TTCCTAAAAATAGTTTTAGAGAT 1251
```

```
Db        459  TTTTTTTGGTTTTTTAATTGAT  481
               ||  |     ||  |     |||  |||
```

```
RESULT 8
ID     ABZ10100  standard; DNA; 8056 BP.
XX
AC     ABZ10100;
XX
DT     16-JAN-2003  (first entry)
XX
DE     Haematopoietic cell proliferation disorder related DNA sequence #240.
XX
KW     Human; haematopoietic cell proliferation disorder; cytostatic;
KW     gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW     cytosine methylation state; gene; ds.
XX
OS     Homo sapiens.
XX
PN     WO200277272-A2.
XX
PD     03-OCT-2002.
XX
PF     26-MAR-2002; 2002WO-EP003401.
XX
PR     26-MAR-2001; 2001US-0278333P.
XX
PA     (EPIG-) EPIGENOMICS AG.
XX
PI     Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI     Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI     Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI     Schwope I, Ziebarth H;
XX
DR     WPI; 2003-018942/01.
XX
PT     Detecting and differentiating between hematopoietic cell proliferative
PT     disorders, comprises contacting a target nucleic acid with a reagent that
PT     distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS     Claim 28; SEQ ID NO 240; 117pp; English.
XX
CC     The present invention describes a method for detecting and
CC     differentiating between haematopoietic cell proliferative disorders
CC     associated with at least 1 gene and/or their regulatory regions in a
CC     subject. The method comprises contacting a target nucleic acid in a
CC     biological sample obtained from the subject with at least 1 reagent,
CC     which distinguishes between methylated and non-methylated CpG
CC     dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC     represent specifically claimed nucleotide sequences from the present
CC     invention. Oligonucleotides from the present invention can be used; for
CC     differentiating between healthy haematopoietic cells and proliferative
CC     disorder haematopoietic cells; for differentiating between acute
CC     lymphocytic-leukaemia and acute myelogenous leukaemia; as probes for
CC     determining the cytosine methylation state and/or single nucleotide
CC     polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC     related sequences and their complements; and as primers for the
CC     amplification of haematopoietic cell proliferation disorder related DNA
CC     sequences. The nucleotide sequences from the present invention can also
CC     be used for detecting a predisposition to, differentiation between
CC     subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC     haematopoietic cell proliferative disorders. The present method enables a
CC     highly specific classification of haematopoietic cell proliferative
CC     disorders allowing for improved and informed treatment of patients
XX
SQ     Sequence.8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;

Query Match          3.2%;   Score 51.8;  DB 7;  Length 8056;
Best Local Similarity 42.3%;  Pred. No. 0.16;
Matches 403; Conservative  0; Mismatches 547; Indels  3; Gaps  2;

Qy       449  ATGATGATCTGAATAAACATACCAGGATAAAGATGTCAAATGAGTGTGACTCCCTTAAAG  508
```

```
Db         96  ATTTTTTATATATTATTAATTATTAATTTTTAATTTTTTTATATTTTTTATTTATTTTTTA  155
               ||||||   |  |     |||  ||  | |   ||   |||  |||

Qy        509  TAGATTAAAGTGTGCATTCTTGTTCCTAAATATGATTTTACTGCTTGAAATTACATT  568
               ||||

Db        156  TTTTTTTATTATTTTTTATTATTTTTTTAATTTTTAATTTTAATTATTATTATATTT  215

Qy        569  TGAGTTGAAGTTTAGAACTAACATAGCATTAATATGAATAATGCATGGAAAATTATTAT  628
               ||||||  ||||

Db        216  TTAAAAATTATTTTAAAATTTTAAAATTTTAAAATTTAAAATTTAAATTTATTTT  275

Qy        629  CCTTTGAAAACTGATTGATAAATATATTCCCCCTCCTTTAGAAACAGTCAAAAGCCACTT  688
               ||||

Db        276  AATAAAAAAAAATTTTTATTTTTTTAAAATTTTTTA-AAAAATAATTAATAAAAA  334

Qy        689  CAAACAAGTTCAAAATAAAGGAGGTAGCAAGTTAGGCGATGGATATATTTTCTTGGC  748
               ||||||

Db        335  TAAATAAAAATAATAATTAAAAAAAAAAAACGAAATAAAATATTTATAATTTAA  394

Qy        749  TTGTTGTATACCCATTGGCCAGGCCTTTATAAGGACTCCCAAAAGCATTTGAAGAATG  808
               |||

Db        395  AATTTAATTTAATAATATATTTATTAATTTATTTTTTTAAAAAATAATAAAAAAAA  454

Qy        809  GCAATATCAAATAAGTGTATGTCCTCTCAAATGAGGAGCATTTTAATTGTTACAATCTATT  868
               |||

Db        455  ATTAATTTTAAAAAATAAATATATTTAAATTAAAAATTAAAATTTATAATATTTATA  514

Qy        869  TGGACGCTCAGGTTATGATATGTTTATGAAAAAATAAGCTTCATTATTCTTATAGCTACA  928
               |||

Db        515  ATATTAAATATTATATTAAAATTAAAATTTTATATTATTATATAACGTTTTATTAA  574

Qy        929  TCCTATTATTCCCTTTAGAAACAAGAATAACAATAAGTTTAATAGTTGCCATACTTAG  988
               |||

Db        575  AATTATTTTATATTTAATTTAAATTAATTTTAAAATTTTAAAATTTTT  634

Qy        989  CATTTATCAGGTCTAATGAAACCAATATTGAATCTCTGATAAATATTTTCTGATGTTACT  1048
               |||

Db        635  TAAAAATAATTTTTAAGATAATATAAATTTATTTAAAAAAAAAAAATATAATATAAAAA  694

Qy       1049  AGCTATGGAAATTAGAACTGGCA--CAACCCTGACATTACTAAGTGAAATGTTAGGAT  1106
               |||

Db        695  AATTTTTTTATATTTTAAAATTTAAAAATAAATAATTTATATATTAATTAAATTTTTAAT  754

Qy       1107  TTTTTCGGCATCGCATGTTAGAATCTCAAAATTCTAAAATTCCTGTTAAATGACTAAGGT  1166
               |||

Db        755  TTTTAAAATTTAATTAAAAAATTATAAAATTATATTTAATTATTTAAAATATAAATTTATA  814

Qy       1167  TTGCTTTATCAATATGAATTCTGAAGGCCAATATCATACCATTAACTATGAAAGCTTTT  1226
               |||

Db        815  AACGATTATTTAAAAATATTTTATTATAAAAAAAATCGTTTTTTTTATTTTTTTTACGTAT  874

Qy       1227  AATTCCTAAAAATAGTTTTAGAGATATTCAAGCAATGCTCTCCTAATCCATACGCAAG  1286
               |||

Db        875  AAAAATTCGTAATTATTTAAATTTAATTTTTCGTGTTTTCGTTAATTAAT  934

Qy       1287  TGTGTTATGACACAAATTCACTAGTCTGGTTTAAAAATGAAATCTTTATATTGACTCGG  1346
               |||

Db        935  TTTTTTAAAAATAATTTTTTAAAAATTTTTTATTTTTTATTATTTTTTAATTCGAAATT  994

Qy       1347  TGTCCCACATATTCCAGTAATTCTCGTTATGAGGAGGACTTGAAATAGCAAATT  1399
               |||

Db        995  TTTTAAAAATAATTAATTTTTTAAAATTTTTTATTATTTTTTAATTTTCGAAATT  1047
```

```
RESULT 9
ID     ABZ10246  standard; DNA; 8056 BP.
XX
AC     ABZ10246;
XX
DT     16-JAN-2003  (first entry)
XX
DE     Haematopoietic cell proliferation disorder related DNA sequence #386.
```

Db 1816 ATAAAATAAAATTTGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATTAATTAAT 1875
Qy 744 TTGGCTTGTGTATACCCATTGGCCAGGGCCTTTATAAGGACTCCCAAAAGCATTTTGAA 803
Db 1876 TATTTTTTATTTTTTTATTTAATATGATAAAAATTTTATAAAAATAAATTAATTAAATTA 1935

Qy 804 GAATGGCAATATCAAATCAAATAGTGTATGTCCTCTCAAATGAGGCATTTTAATTGTTACAAT 863
Db 1936 TAATATAAATAAAAAAATAAATATTAATTTTTAAAATAAAATATTAATTAATTTTATTTA 1995

Qy 864 CTATTTGGACGCTCAGGTTATGATATGTTTATGATATGTTTATGAAAAATAAGCTTCATTATTTCTTATAG 923
Db 1996 AAATTTAATAAAATTTTAAATATTTAAAATATATTTTTAAATATTTTTATTT 2055

Qy 924 CTACATCCTATTATT--CCCTTTAGAAACAAGAATAACAATAAGTTTAATAGTTGCCA 981
Db 2056 TTAAAAAATTTTATTTAAAAATATTTTAAATATTTATAAAATAATAATATAAATTAATTAAT 2115

Qy 982 TACTTAGCATTTATCAGGTCTCAATGAAACCAATATTGAATCTCTGATAAATATTTTCTGA 1041
Db 2116 ATAATTTTAATAATTTTATTTATTTATTAAAATAAAATAAAAATAAAAAAAATTATAA 2175

Qy 1042 TGTTACTAGCTATGGGAAATTAGAACTGGACACAACCCTGACATTACTAAGTGGAAATGTT 1101
Db 2176 TTTTATTAAAAATTAAAAAAATGAAAAAAAAAATTAATTTTA 2235

Qy 1102 AGGATTTTCGGCATGCATGTAGAATCTCTAAAATTTAAACATTCCTGTTAAAATGACT 1161
Db 2236 ATAAAAATT--------TATTTTTTAATATTTAAAATTAAAAAATTTAATTTTT 2287

Qy 1162 AAGGTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCATACCATTAACTATGAAAG 1221
Db 2288 TAATTATTATTTTATTAAAAATAATTTTTGTATTTAAATAAATTTTATTTAAT 2347

Qy 1222 CTTTTAATTCCTAAAAATAGTTTAGAGATATTCAAGCAA 1261
Db 2348 TTTTAAATATAATTTTTTTATTTTTATAAATAATAATTATTAA 2387

RESULT 10

ID   ABK28407 standard; DNA; 9504 BP.
ABK28407/c
XX
AC   ABK28407;
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA transcription associated genomic DNA #141.
XX
KW   DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW   PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW   single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW   viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW   immunological disorder; Werner syndrome; developmental disorder;
KW   psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW   neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW   myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW   angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW   polyglutamine disorder; solid tumour.
XX
OS   Unidentified.
XX
PN   WO200192565-A2.
XX
PD   06-DEC-2001.
XX
PF   06-APR-2001; 2001WO-EP003973.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX

XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-EP003401.
XX
PR   26-MAR-2001; 2001US-0278333P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI   Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI   Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI   Schwope I, Ziebarth H;
XX
DR   WPI; 2003-018942/01.
XX
PT   Detecting and differentiating between hematopoietic cell proliferative
PT   disorders, comprises contacting a target nucleic acid with a reagent that
PT   distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS   Claim 28; SEQ ID NO 386; 117pp; English.
XX
CC   The present invention describes a method for detecting and
CC   differentiating between haematopoietic cell proliferative disorders
CC   associated with at least 1 gene and/or their regulatory regions in a
CC   subject. The method comprises contacting a target nucleic acid in a
CC   biological sample obtained from the subject with at least 1 reagent,
CC   which distinguishes between methylated and non-methylated CpG
CC   dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC   represent specifically claimed nucleotide sequences from the present
CC   invention. Oligonucleotides from the present invention can be used: for
CC   differentiating between healthy haematopoietic cells and proliferative
CC   disorder haematopoietic cells; for differentiating between acute
CC   lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC   determining the cytosine methylation state and/or single nucleotide
CC   polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC   related sequences and their complements; and as primers for the
CC   amplification of haematopoietic cell proliferation disorder related DNA
CC   sequences. The nucleotide sequences from the present invention can also
CC   be used for detecting a predisposition to, differentiation between
CC   subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC   haematopoietic cell proliferative disorders. The present method enables a
CC   highly specific classification of haematopoietic cell proliferative
CC   disorders allowing for improved and informed treatment of patients
XX
SQ   Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

Query Match   3.2%; Score 51.2; DB 7; Length 8056;
Best Local Similarity   43.7%; Pred. No. 0.21;
Matches 332; Conservative   0; Mismatches 418; Indels 10; Gaps 2;

Qy 504 TAAAGTAGATAAAGTGTGCATTCTTGTTCCTAAAAATAATGATTTTACTGCTGAAATT 563
Db 1636 TAAAATTAAAAAAAAATTAATTATTTTAATTAATGATAAAATTTTATTTTTATTAAAATTATA 1695

Qy 564 ACATTTGAGTGAAGTTAGAAACTAGAAAACTAACATAGCATTAAATGAATGAAATT 623
Db 1696 AAAATTAATTAATTAATAATAATATATATTTATAATTTTATTAATTTTAAAAAATTAAAATA 1755

Qy 624 ATTATCCTTTGAAAACTGATTGATAAGATGATAAAATAATATTCCCCCTCCTTTAGAAACAGTCAAAAGC 683
Db 1756 AAAAAATTTTTTAAAATTTAAAAAAATTAAAAATTAATATGTTATTATATTTTAATTTAATA 1815

Qy 684 CACTTCAAACAAGTTCAAAATAAAGGAAGGTAGCAAGTTAGGCGATGGATTATATTTTC 743

PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-090046/12.
XX
PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT cancer.
XX
PS Claim 1; SEQ ID NO 281; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequence ABK8127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office
XX
SQ Sequence 9504 BP; 2528 A; 138 C; 1853 G; 4985 T; 0 U; 0 Other;

Query Match   3.1%; Score 50.6; DB 6; Length 9504;
Best Local Similarity   53.9%; Pred. No. 0.3;
Matches 104; Conservative   0; Mismatches   89; Indels   0; Gaps   0;

Qy   846 ATTTTTAATTGTTACAATCTATTTGGACGCTCAGGTTGATATGTTTATGAAAAATAAG 905
            |||||||||| |||| ||||||| ||||| ||| ||||| | ||||||||||||
Db   7045 ATTTTTAATTAATTAACTATTTTTTTATCACAACTAATTAATAATATTCCATAAAATTAAAAATC 6986

Qy   906 CTTCATTATTTCTTTATAGCTACACATCCTATTATTCCCTTTAGAAACAAGAATAACAATAA 965
            ||| ||||||||||||| ||| ||| ||||||||| ||||||||||||||||||||||||
Db   6985 ATACATATTTAACAATTATTCTTAAAACTTCATATTTATAATTTAAAAAAAAATTAAAAAAA 6926

Qy   966 GTTTTATAGTTGCCATACTTAGCATTTATCAGGTCTAATGAAACCAATATTGAATCTCT 1025
            ||| | |||| |||| |||||| ||| |||||||||||||||||||| ||||||
Db   6925 TTTCAAATAACTTCTATTTTAAAATAATCTATATTTAAAATAATAATAAAACATCTACTTA 6866

Qy   1026 GATAAATATTTTC 1038
            ||||||||||||
Db   6865 AAAAAAATTTTCC 6853

RESULT 11
ABQ67093/c
ID ABQ67093 standard; DNA; 83391 BP.
XX
AC ABQ67093;
XX
DT 28-AUG-2002 (first entry)
XX
DE Human angiogenesis associated polynucleotide SEQ ID NO 123.
XX
KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

KW antiarteriosclerotic; ds.
XX
OS Homo sapiens.
XX
PN WO20246454-A2.
XX
PD 13-JUN-2002.
XX
PF 06-DEC-2001; 2001WO-EP014320.
XX
PR 06-DEC-2000; 2000DE-01061338.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Schacht O;
XX
DR WPI; 2002-500450/53.
XX
PT New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.
XX
PS Claim 1; SEQ ID NO 123; 41pp + Sequence Listing; German.
XX
CC The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 83391 BP; 25113 A; 665 C; 16761 G; 40835 T; 0 U; 17 Other;

Query Match   3.1%; Score 50.4; DB 6; Length 83391;
Best Local Similarity   43.8%; Pred. No. 0.53;
Matches 266; Conservative   0; Mismatches 341; Indels   1; Gaps   1;

Qy   428 CCCTTAAGGAACTTATAATTAATTAATGATGATCTGAATAAACATACCAGGATAAAGATGTCAA 487
            ||||||| |||| || |||||||||||||||
Db   19871 CTCTTAATAAACTAAAAATTAAATAATTAATAATTTTAAATATCTATCTTAAACCAATAAA 19812

Qy   488 ATGAGTGTGACTCCCTTAAAGTAGATTAAAGTGCATTCTTTGTTTCCTAAAATATGAT 547
            ||| |||||||| || ||||
Db   19811 CAACATATTTAACAATTATTATTAAAAAAACATTTACCATTACCAACAAATTAAAAAAAATA 19752

Qy   548 TTTACTGCTTGAAATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTAATATGAA 607
            ||| ||||| || || |
Db   19751 CTTATAATCATAAATTTTAATTAACTATTTTAAAAACTAATAATAATTTATAAAC-TATACCATTAATCTAA 19633

Qy   608 TAATGCATGAAAATTATTATCCTTTGAAAACAAGTTTCAAAATAAAAGAAAGGAAGGAAGTTAGGC 667
            | | ||| |||||||| ||| ||||||||
Db   19632 ACAAACACTAAAAATATAAAAATTAAAACTAAAATCATTAAAAATCATTACAAAACTT 19573

Qy   728 GATGGATTATATTTCTTGGCTGTGTTGTATACCCATTATTAAAAATAAAATAACCTTCTAATACACCAAT 787
            ||||||| ||||
Db   19572 ATTAAAAATAACTAATTTCAAATAACATTTAAAAATAAATAACCTTCTAATACACCAAT 19513

Qy   788 CCAAAAGCATTTTGAAGAATGGCAATATCAAATATCAAATGTATGTCCTCTCAAATGAGGCAT 847
            || ||||||||||||
Db   19512 AATAATCCATTAAAAAAAAACCAAAAAAAAATTATATTTTCATTAACAACCAAAAA 19453

Qy   848 TTTTAATTGTTACAATCTATTTGGACGCTCAGGTTATGATATGTTTATGAAAAATAAGCT 907
            || |||| ||| ||||| |||| |||

```
Db  19452 ATTAAATTTATATATATAATTTATATACAACTATATATAATAATATATATATAATTAAATT 19393

Qy    908 TCATTATTTCTTATAGCTACATCCTATTATTCCCTTTTGAAACAAGAATAACAATAAGT 967
                ||||||| |||   ||||||||||||||||||||||  ||||||||| |||||||||
Db  19392 CTATATTTATATTATACAATATAATAAAACTTTTTTAAAAAATATATAAAATCTATAA 19333

Qy    968 TTTAATAGTTGCCATACTTAGCCATTTATCAGTCTCATGAAACCAATATTGAATCTCTGA 1027
                |||||||      ||||||||| ||||  ||
Db  19332 AAAAAAATATAAAAAACTATAAAAATCATTATTAAAAACTCGACTACACATAAACTATAC 19273

Qy   1028 TAAATATT 1035
                |||||
Db  19272 ATCATATT 19265
```

RESULT 12
ABL56203/c
ID ABL56203 standard; DNA; 32392 BP.
XX
AC ABL56203;
XX
DT 01-JUL-2002 (first entry)
XX
DE AmEPV genome fragment#5.
XX
KW AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
KW genetic deficiency disorder; ds.
XX
OS Amsacta moorei entomopoxvirus.
XX
PN WO200211526-A2.
XX
PD 14-FEB-2002.
XX
PF 10-AUG-2001; 2001WO-US025287.
XX
PR 10-AUG-2000; 2000US-0224479P.
PR 14-SEP-2000; 2000US-00662254.
XX
PA (UYFL) UNIV FLORIDA.
XX
PI Moyer RW, Li Y, Bawden AL;
XX
DR WPI; 2002-227161/28.
XX
PT Novel recombinant entomopox virus vector useful for delivering
PT polynucleotide encoding protein to vertebrate cell, comprises
PT polynucleotide encoding protein operably linked with heterologous
PT promoter sequence.
XX
PS Disclosure; Page 226-242; 326pp; English.
XX
CC The invention relates to a recombinant entomopox virus (EPV) vector,
CC comprising a polynucleotide encoding a protein operably linked with a
CC heterologous promoter sequence. The invention also concerns methods for
CC providing gene therapy for genetic deficiency disorders. Vectors of the
CC invention are useful for delivering a polynucleotide encoding a protein
CC to a vertebrate cell preferably a mammalian cell, such as a human cell.
CC The vector is introduced into the vertebrate cell by infection in a viral
CC particle, or by transfection, transduction, or injection either in vitro
CC or in vivo. The vector is useful for the delivery and expression of
CC biologically useful proteins in gene therapy protocols, and for
CC delivering large DNA segments for engineering of vertebrate cells.
CC Polynucleotides of the invention have applications in techniques such as
CC their use as insertion sites for foreign genes of interest, hybridisation
CC probes, for chromosome and gene mapping, in PCR technologies, and in the
CC production of sense or antisense nucleic acids. Vectors of the invention
CC provide for stable integration and expression of heterologous DNA in host
CC cells, and are adapted for accepting large heterologous polynucleotide
CC inserts which can be delivered in an infected or transformed cell and
CC expressed in a stable fraction. The current sequence represents a
CC fragment of the genome of the genus B entomopoxvirus from amsacta moorei
CC (AmEPV)

XX
SQ Sequence 32392 BP; 13748 A; 2577 C; 2550 G; 13517 T; 0 U; 0 Other;

Query Match   3.1%; Score 49.8; DB 6; Length 32392;
Best Local Similarity 44.9%; Pred. No. 0.58;
Matches 391; Conservative   0; Mismatches 472; Indels   8; Gaps   5;

```
Qy    391 AAAAGAGAACTGTGTAATTGAAATGACGTGGGCTGCACCCTTAAGGAACTTATAATTAAT 450
              |||   ||||| |||||| |||||||||||||||||    ||||||||||||||
Db   1327 AAATCAGTATATAGTAAAACAGAAGAAGGGAGTAATTCTGTTATTAACGTGATATCTGAT 1268

Qy    451 GATGATCTGAATAAACATACCAGGATAAAGTGTCAAATGAGTGTGACTCCCTTAAAGTA 510
              |||| ||||    ||||||||||||||||||||||||||||||
Db   1267 AATAAATTAAAAAAATAGTTTAGCAGAAGATATATCTTTAAAAAGGATGTATAAAAGAA 1208

Qy    511 GATTAAAGTGTGCATTCTTTGTTCTCAAATATGATTTTACTGCTTGAAATTACATTTG 570
              ||||||
Db   1207 GATTATAATATAGATT-TATATGAGAAAAAAATATATGTAAAAAATAATAGAAAAATAA 1149

Qy    571 AGTTGAAGTTTAGAAACTAACATAGCCATTAATAATGAATAATGCATGGAAAATTATTATCC 630
              |||||||||| |||  ||  ||||  ||
Db   1148 ATATTATATCAAAAAATTATTATCTTTGATTAT-AATAATAAAAAAATAATAATATATAG 1090

Qy    631 TTTGAAAACTGATTGATAAATATATTCCCCTCCTTAGAAACAGTCAAAAGCCACTTCA 690
              ||||||||||| ||
Db   1089 TGTGTGGTAAAATTGATGGAATTATTAACGATGTATTAATTGAAAGCAAAATCGAAGAA 1030

Qy    691 AACAAGTTCAAAATAAAGGAAGGTAGC----AAGTTAGGCGATGGATTATATTTTCTTG 746
              ||| ||| ||    |||||||||||
Db   1029 AAAGACTATTCAAATACATTCCTATATATGAAAAAGTTCAATTAGAAAAATTATATTTTT 970

Qy    747 GCTTGTTGTATACCCATTGGCCAGGGCCTTTATAAGGACTCCCAAAAGCATTTTGAAGAA 806
              ||| |||| ||       |||||| ||   |||| |||  || |
Db    969 TATTGAATTATAAAAGAGTAAATTAATTCAACATTATGATAACGAAATAGATGTAATAG 910

Qy    807 TGGCAATATCAAATAGTGTATGTCCCTCTCAAATGAGGCATTTTAATTGTTACAATCTA 866
              ||||  ||     ||||| ||||  |  |||| ||  |||  || || |||
Db    909 ATTATATCCCGACGAACAATTATTTTCTGAAATATGTAATAATGTAATATCATATTTTA 850

Qy    867 TTTGGACGCTCAGGTTATGATATGTTTATGATAAAAATAAGTTCATTATTTCTTATAGCTA 926
              ||| |  |||||| ||| || || |||||
Db    849 TAAAAATGTTTA-GTAATAA-TATATTTAAAATATAAATATAGATTATATAATAAATTTAATTAA 791

Qy    927 CATCCTATTATTCCCTTTAGAAACAAGAGAATAACAATAAGTTTAATAGTTGCCATACTT 986
              ||
Db    790 TATTATATTCTATAATAATAATTTTTGGATTAATATCGATATAATTATAATAATATATA 731

Qy    987 AGCATTTATCAGGTCTAATGAAACCAATATTGAATCTCTGATAAATATTTTCTGATGTTA 1046
              |||  ||||  ||||   ||  ||||||  || | | ||||
Db    730 TATTAGTATTAATAAAATTTATATAAAAATATTAGATATATCTTATCAAATTTATAAAAAATAT 671

Qy   1047 CTAGCTATGGGAAATTAGAACTGGCACAACCCTGACATTACTAAGTGGAAATGTTAGGAT 1106
              |||   |  ||||||||| ||  |||  | |   ||| |||
Db    670 TACCATTATATATATTATCATTATATTGATTATATTTACACAATTATTATATTTTTTTTA- 612

Qy   1107 TTTTCGGCATCGCATGTTAGAATCTCTAAAATTTAAACATTCCTGTAAATGACTAAGGT 1166
              |||  ||   | ||||||      |||
Db    611 TCATATATATATTTTTATCTATTGTTATTGTTATTACTAACAATTCTCTATTATCT 552

Qy   1167 TTGCTTTTATCAATATGAATTCTGAAGGCCAATATCATACCATTAACTAACATGAAAGCTTTT 1226
              |||||  ||  |||||  || ||
Db    551 ATATTTTTTTATTTTATATTCATATTCATAATCAATAATTTATTTATTAATTATATTATTACTTA 492

Qy   1227 AATTCCTAAAAAAATAGTTTTAGAGATATTCAA 1257
              ||| |
Db    491 TATAATAATATATCATTTATATATTTTATA 461
```

RESULT 13
AAL15210
ID AAL15210 standard; cDNA; 883 BP.
XX
AC AAL15210;

```
Qy    800  TGAAGAATGGCAATATCAAATAAGTGTATGTCCTCCAAATGAGGCATTTTAATTGTTA  859
           |||| | |||| |||  |||| ||  ||||||||  ||||||||||||| |||||||||
Db    620  TTTATNNAAANAAAAAAAAANANTTGNTTAAAATTTNAAATTTTTTTTTTTNATTAAAA  679

Qy    860  CAATCTATTTGGACGCTCAGGTTATGATATGTTTATGATATGTTTATGAAAAATAAGCTTCATTATTTCTT  919
           ||||| |||    ||||||| | ||||||| | |||||||| | ||||||||  |||||||||
Db    680  AATTNAANTTATTAANTTTATTTTTTTAAAAATTTTTTAAAAAAAAATTTTGGTTTAATRANT  739

Qy    920  ATAGCTACATCCTATTATTCCCTTTAGAAACAAGAATAACAATAAGTTTTAATAGTT  977
           |||||   | ||| |||  ||||| ||||||  |   ||||||||  |||||||||
Db    740  TANAAAANNTTTAATTTTNNNTTTTTTTANAAAAATTNANTATNNTTANAATTAGTT  797
```

RESULT 14

ABZ10246/c

XX

ID   ABZ10246 standard; DNA; 8056 BP.

XX

AC   ABZ10246;

XX

DT   16-JAN-2003 (first entry)

XX

DE   Haematopoietic cell proliferation disorder related DNA sequence #386.

XX

KW   Human; haematopoietic cell proliferation disorder; cytostatic;

KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;

KW   cytosine methylation state; gene; ds.

XX

OS   Homo sapiens.

XX

PN   WO200277272-A2.

XX

PD   03-OCT-2002.

XX

PF   26-MAR-2002; 2002WO-EP003401.

XX

PR   26-MAR-2001; 2001US-0278333P.

XX

PA   (EPIG-) EPIGENOMICS AG.

XX

PI   Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;

PI   Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;

PI   Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;

PI   Schwope I, Ziebarth H;

XX

DR   WPI; 2003-018942/01.

XX

PT   Detecting and differentiating between hematopoietic cell proliferative

PT   disorders, comprises contacting a target nucleic acid with a reagent that

PT   distinguishes between methylated and non-methylated CpG dinucleotides.

XX

PS   Claim 28; SEQ ID NO 386; 117pp; English.

XX

CC   The present invention describes a method for detecting and

CC   differentiating between haematopoietic cell proliferative disorders

CC   associated with at least 1 gene and/or their regulatory regions in a

CC   subject. The method comprises contacting a target nucleic acid in a

CC   biological sample obtained from the subject with at least 1 reagent,

CC   which distinguishes between methylated and non-methylated CpG

CC   dinucleotides within the target nucleic acid. AB209861 to ABZ11118

CC   represent specifically claimed nucleotide sequences from the present

CC   invention. Oligonucleotides from the present invention can be used: for

CC   differentiating between healthy haematopoietic cells and proliferative

CC   disorder haematopoietic cells; for differentiating between acute

CC   lymphocytic leukaemia and acute myelogenous leukaemia; as probes for

CC   determining the cytosine methylation state and/or single nucleotide

CC   polymorphisms (SNPs) of haematopoietic cell proliferation disorder

CC   related sequences and their complements; and as primers for the

CC   amplification of haematopoietic cell proliferation disorder related DNA

CC   sequences. The nucleotide sequences from the present invention can also

CC   be used for detecting a predisposition to, differentiation between

CC   subclasses, diagnosis, prognosis, treatment and/or monitoring of

CC   haematopoietic cell proliferative disorders. The present method enables a

CC   highly specific classification of haematopoietic cell proliferative

---

XX

DT   07-DEC-2001 (first entry)

XX

DE   Human breast cancer expressed polynucleotide 7667.

XX

KW   Human; breast cancer; cell marker; cytostatic; ss.

XX

OS   Homo sapiens.

XX

PN   WO200151628-A2.

XX

PD   19-JUL-2001.

XX

PF   10-JAN-2001; 2001WO-US000798.

XX

PR   14-JAN-2000; 2000US-0176077P.

PR   14-MAR-2000; 2000US-0189167P.

PR   24-MAR-2000; 2000US-0192099P.

PR   29-MAR-2000; 2000US-0193480P.

PR   15-MAY-2000; 2000US-0205230P.

PR   09-JUN-2000; 2000US-0211315P.

PR   25-JUL-2000; 2000US-0220534P.

XX

PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX

PI   Lillie J, Xu Y, Wang Y, Steinmann K;

XX

DR   WPI; 2001-451856/48.

XX

PT   New peptide useful as a marker for the diagnosis of breast cancer.

XX

PS   Claim 1; Page 1378; 3695pp; English.

XX

CC   The invention relates to human breast cancer expressed polynucleotides

CC   (AAL07544-AAL26789) and methods of assessing whether a patient is

CC   afflicted with breast cancer by examining the correlation between the

CC   expression of certain markers and the cancerous state of breast cells.

CC   The polynucleotides and encoded polypeptides are potential markers for

CC   detecting, diagnosing, monitoring, characterising treating and

CC   potentially preventing breast cancer. The polynucleotides and encoded

CC   polypeptides are also useful for isolating compounds with cytostatic

CC   activity

XX

SQ   Sequence 883 BP; 322 A; 18 C; 23 G; 396 T; 0 U; 124 Other;

Query Match   3.0%; Score 48.8; DB 4; Length 883;

Best Local Similarity 38.3%; Pred. No. 0.45;

Matches 206; Conservative   0; Mismatches 332; Indels   0; Gaps   0;

```
Qy    440  TTATAATTAATGATGATCTGAATAAACATACCAGGATAAAGATGTCAAATGAGTGTGACT  499
           |||| | || ||  |||||| ||| ||||| |||| || ||| ||||||||  |||||
Db    260  TTTTTATATAAAAANAATTTTTTTTTTAATTATAAAATTANAAATANGTTTAAANTTATT  319

Qy    500  CCCTAAAGTAGATTAAAGTGTGCATTCTTCTTTGTTTCCTAAAATATGATTTTACTGCTTGA  559
           ||| ||||| |  |||||| | ||   |||| |  ||||||| |||| |
Db    320  TTATTNATTAAANTAAAATTTTNTNTNNANTTTNTAAATTNTNTNTNNNATAANTATAAA  379

Qy    560  AATTACATTTGAGTTGAAGTTTAGAAAACTAACACATAGCATTAATAATGATAATGCATGGAA  619
           |||||| || || ||  ||  ||||||||  ||| |||| | ||||||||||||||||
Db    380  ATTATTTTTTTNAAAATTATTTTATTNTNTNTNTTTTAATTTATAGAANTAAAAAAATAA  439

Qy    620  AATTATTATCCTTGAAAACTGATTGATAAATATATTCCCCTCCTTTAGAAACAGTCAA  679
           |||||  |  |||| || |||||||||| || |  ||||||||| |||
Db    440  ATTTTTATNTNNAANNNAAAAAAATTNTNTTNAATTNAAGGTTTTAAAAANANTTN  499

Qy    680  AAGCCACTTCAAACAAGTTCAAAATAAAGGAAGGTAGCAAGTTAGGCGATGGATTATAT  739
           |||  |    |||||||| ||||  || || |  |||||||||||||||
Db    500  AAAAAATTTTTTTNTTTNAGAAAAANANAAAATAAAATTTTTATAAANTATAAAAAA  559

Qy    740  TTTCTTGGCTTGTGTATACCCATTGGCCAGGGCCTTTATAAGGACTCCCAAAGCATTT  799
           |||  |    || |||||||  ||||||||| || || |||||||||||
Db    560  AAAAANNTTTTTTTTTTAAATATNTTNATAAAAAANNTNNTTTTTTATTTAAAAANTTTT  619
```

CC   disorders allowing for improved and informed treatment of patients
XX
SQ   Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

Query Match   3.0%;   Score 48.8; DB 7; Length 8056;
Best Local Similarity 44.0%;   Pred. No. 0.72;
Matches 297; Conservative 0; Mismatches 374; Indels 4; Gaps 2;

Qy   438 ACTTATAATTAATGATGATCTGAATAAACATACCAGGATAAAGATGTCAAATGAGTGTGA 497
          ||| ||||||| |||| |||
Db   2148 ATTTTTAATAAAATAAAAATATTATTAAAATAATTATAATTATAATTATTATTATTAT 2089

Qy   498 CTCCCTAAAGTAGATTAAAGTGTGCATTCTTTGTTCCTAAAATAAAAAATATTTTACTGCTT 557
          ||||||| ||||| ||||||
Db   2088 AATATTTAAAATATTTTTAATAAAAATTTTTAATAAAAATAATAAAAAATATTAATAATATTT 2029

Qy   558 GAAATTACATTGAGTTGAAGTTTAGAAA---CTAACATAGCATTAAATGAATAATGCA 614
          ||||| |||| ||||||| |||||
Db   2028 TATTTTAAATATATTAAAATTTTTATTAAATTTAAATAAAAATTAATAATAATAATTTTA 1969

Qy   615 TGGAAAATTATTATCCTTTGAAAACTGATTGATAAATATATTCCCCCTCCTTTAGAAACA 674
          ||||||||||||| |||||
Db   1968 TTTTAAAAAATTATATTTTTTTTTATTTATTTATATTAATTTAAATTATTTTTTTTTATAAA 1909

Qy   675 GTCAAAAGCCACTTCAAACAAGTTCAAAAATAAAGGAAGGTAGCAAGTTAGGCGGATGGAT 734
          ||||||||||| ||| |||
Db   1908 ATTTTTATTTTTATAATAAAAAAAATAAAATAATTAATTAAATTAATTTTTTTTTTTTTT 1849

Qy   735 TATATTTTCTTGGCTTGTTGTATACCCATTGGCCAGGGCCTTTATAAGGACTCCCAAAAG 794
          ||| ||||||
Db   1848 TTTTTTTTTTTTAAAACAAAATTTTATTTTATTTTAAAATAAAATAATAACATAT 1789

Qy   795 CATTTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCTCAAATGAGGCATTTTTAAT 854
          |||||||| ||| ||
Db   1788 TAATTTTTAATTTTTAATTTTTAAAATTTTTTTATTTAATTTTTTAAATAATTAAAAT 1729

Qy   855 TGTTACAATCTATTTGGACGCTCAGGTTATGATATGTTTATGAAAAATAAGCTTCATTAT 914
          |||| ||| |||||
Db   1728 TATAAATATATTTATTATTA-TATTAAATAAATTAATTTTTATAATTTTTAATAAAAATAAAATTTT 1670

Qy   915 TTCTTATAGCTACATCCTATTATTCCCTTTAGAACAAGAATAACAATAAGTTTTAATA 974
          |||||| ||| ||||
Db   1669 ATCATATTAATTAATAAAATATTTTTTTTTAATTTTTCAATTCATTTATTTCAATTTTTTAAA 1610

Qy   975 GTTGCCATACTTAGCATTTATCAGGTCTCAATGAAATCCAATATTGAAATCTCTGATAAATAT 1034
          |||||||| ||| ||
Db   1609 ATATTAAAATTATAAATATAAAATAATTAAAATTATAATTTTTAAAAAATAATTTTTAAAAAAAT 1550

Qy   1035 TTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACATTACTAAGTGG 1094
          ||||| ||| ||
Db   1549 AATTTTATTTAAAAATTTTTTTTAACATATTTTTATATTTTTATATATATTTTTATAATTTTTA 1490

Qy   1095 AAATGTTAGGATTTT 1109
          ||||||
Db   1489 ATAATTAATTATTTT 1475

RESULT 15
ABL32891/c
ID   ABL32891 standard; DNA; 6131 BP.
XX
AC   ABL32891;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 864.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO20020928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 864; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6131 BP; 1970 A; 39 C; 1013 G; 3109 T; 0 U; 0 Other;

Query Match   3.0%;   Score 48.6; DB 6; Length 6131;
Best Local Similarity 48.1%;   Pred. No. 0.75;
Matches 169; Conservative 0; Mismatches 179; Indels 3; Gaps 1;

Qy   888 ATGTTTATGAAAAATAAGCTTCATTATTTCTTATAGCTACATCCTATTATTCCCTTTTAG 947
          ||| ||| ||| ||||
Db   718 ATTATCATAAAAAATATATTACTTTTTACTTAAAAAATAAAAATAAACCAAAATATATTATTATAA 659

Qy   948 AAACAAGAATAACAATAAGTTTAATAGTTGCCATACTTAGCATTTATCAGGTCTAATGA 1007
          |||| ||| ||| ||||
Db   658 AAACATTAAAAAAATATAAAATACATAATCAATAAAACCATCAAATCAAATCAAATAATTA 599

Qy   1008 AACCAATATTGAATCCTCTGATAAATATTTCTGATGTGATGTTACTAGCTATGGGAAATTAGAAC 1067
          ||| ||| ||| |||
Db   598 AATTTATTTCAATATAATATTTAACTCTTCCTAAATAAAATCATTTATCATTTCAAAC 539

Qy   1068 TGGCACAACCCTGACATTACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGA 1127
          |||||| ||| ||||
Db   538 TAATATTATCATCAACCTTAAAATTCTTATAAACAAATATTAAATACAATTAATTAAATTA 479

Qy   1128 ATCTCTAAAATTTAAACATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATATGAATT 1187
          |||||| ||| ||||
Db   478 AA--TATTATATAAAATTAAAAACCCAATTAAAATATACAATATTTATAAACAATTAAATTA 422

Qy   1188 CTGAAGGCCAAATATCATACCATTAACTATGAAAGCTTTTAAATTCCTAAAAA 1238
          |||||| ||| ||||
Db   421 TTCAATAAAAATTTTTTTCCTATAATTATTATATAACATATATTTCTTAAAAA 371

Search completed: April 25, 2004, 08:46:03
Job time : 434.176 secs

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: April 24, 2004, 23:54:27 ; Search time 4263.17 Seconds
(without alignments)
16460.143 Million cell updates/sec

Title: US-10-051-835-15
Perfect score: 1619
Sequence: 1 ttaatctacatatccttacgt.........atttatttaaaaatttaaaa 1619

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pri:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Match	Query Length	DB	ID	Description
1		1417.4	87.1	5900	9	HSM805990	BX537884 Homo sapi
2	c	1415	87.4	157042	9	AC024061	AC024061 Homo sapi
3		1414.2	87.4	2150	9	BC058026	BC058026 Homo sapi
4		1412.6	87.3	5292	9	HSM808902	BX648751 Homo sapi
5		1412.6	87.3	6104	9	HSM808720	BX648571 Homo sapi
6	c	1411.8	87.2	176626	9	AC066614	AC066614 Homo sapi
7		1410.2	87.1	2455	9	AK123805	AK123805 Homo sapi
8	c	1408.6	87.0	3493	9	AK125130	AK125130 Homo sapi
9	c	475	29.3	580	6	AX332151	AX332151 Sequence
10	c	475	29.3	580	6	AX334874	AX334874 Sequence
11	c	475	29.3	580	6	AX335287	AX335287 Sequence
12	c	475	29.3	580	6	AX335859	AX335859 Sequence
13	c	118.2	7.3	231508	10	AC111087	AC111087 Mus muscu
14		118.2	7.3	256138	2	AC116714	AC116714 Mus muscu
15	c	96.2	5.9	202314	2	AC132694	AC132694 Rattus no
16	c	96.2	5.9	271288	2	AC105617	AC105617 Rattus no
17	c	68.8	4.2	250029	2	AE014839	AE014839 Plasmodiu
18		66.6	4.1	1141	6	AX083744	AX083744 Sequence
19		62.6	3.9	111861	9	AC069435	AC069435 Homo sapi
20	c	62.6	3.9	148075	9	AC110014	AC110014 Homo sapi
21	c	60.4	3.7	253001	3	AE014834	AE014834 Plasmodiu
22	c	60	3.7	145992	9	AC134919	AC134919 Homo sapi
23	c	59.8	3.7	7218	6	I66494	I66494 Sequence 14
24	c	59.6	3.7	151900	9	AC107419	AC107419 Homo sapi
25	c	58.8	3.6	212046	10	BX005167	BX005167 Mouse DNA
26		58.8	3.6	316399	2	AC128737	AC128737 Mus muscu
27		58.4	3.6	15421	3	PFCOMPIRA	X95275 P.falciparu
28		58.4	3.6	175559	2	AC145101	AC145101 Homo sapi
29		56.6	3.5	192929	3	AC005505	AC005505 Plasmodiu
30		56.6	3.5	250713	3	AE014850	AE014850 Plasmodiu
31		56.6	3.5	256172	2	AC005139	AC005139 Plasmodiu
32	c	56.4	3.5	213591	2	AC141759	AC141759 Homo sapi
33	c	56.2	3.5	172816	9	AC093899	AC093899 Homo sapi
34	c	55.8	3.5	49306	9	AC115606	AC115606 Dictyoste
35		55.8	3.4	169494	9	AL450083	AL450083 Human DNA
36		55.4	3.4	14001	3	PFCOMPIRB	X95276 P.falciparu
37		55.2	3.4	1141	6	AX083744	AX083744 Sequence
38	c	55	3.4	186165	9	AC112187	AC112187 Homo sapi
39		55	3.4	257757	3	AE014837	AE014837 Plasmodiu
40		54.8	3.4	133877	9	AC120883	AC120883 Homo sapi
41		54.6	3.4	14867	3	AE001398	AE001398 Plasmodiu
42		54.4	3.4	194074	9	AC090095	AC090095 Homo sapi
43	c	54.4	3.4	199810	9	AC018443	AC018443 Homo sapi
44	c	53.8	3.3	101739	9	AL591400	AL591400 Human DNA
45	c	53.8	3.3	122364	9	AC007324	AC007324 Homo sapi

ALIGNMENTS

RESULT 1
HSM805990
LOCUS   HSM805990   5900 bp   mRNA   linear   PRI 17-JUN-2003
DEFINITION Homo sapiens mRNA; cDNA DKFZp686F07114 (from clone DKFZp686F07114);
            complete cds.
ACCESSION   BX537884
VERSION   BX537884.1 GI:31873891
KEYWORDS
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 5900)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.

Pred. No. is the number of results predicted by chance to have a

TITLE   Direct Submission
JOURNAL   Submitted (17-JUN-2003) MIPS., Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the
            Heinrich-Heine-University, Duesseldorf/Germany) within the cDNA
            sequencing consortium of the German Genome Project. This clone
            (DKFZp686F07114) is available at the RZPD in Berlin. Please contact
            the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES   Location/Qualifiers
     source   1..5900
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686F07114"
                     /map="15q21.3"
                     /tissue_type="human fetal kidney"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="fetal"
     gene   1..5900
                     /gene="DKFZp686F07114"
     CDS   43..3351
                     /gene="DKFZp686F07114"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD97880.1"
                     /db_xref="GI:31873892"
                     /translation="MRTSLQAVALWGQKAPPHSITAIMITDQRTIVTGSQEGQPCLW
                     NLSHELKISAKELLFGHSASVTCLARARDFSKQPYIVSAARNGEMCVWNVTNGQCVEK
                     ATLPYRHIAICYHCSFRMTGEWLLCCGEYDVLIIDAKTLAVVHSFRSQFPDWIN
                     CMCIVHSMRIQEDSLLVVSVAGGLKVWDLSSSINSIQEKQDVYEKESKFLESLNCQTI
                     RPCTYTERLLIVVFSKCWKVPYDYCDFSLLLTEVSRNGQFFAGGEVIAAHRILIWTEDG
                     HSTIYQLINSGLSKSIYPADGRVLKETIYLHLLCSTSVQENKEQSRPFVMGYMNERKE
                     PFYKVLFSGEVSGRITLMHIPDVPVSKFDGSPREIPVTATWTLQDNFDKHDTMSQSII
                     DYPSGLKDGAGTAVVTSSEYIPSLDKLICGCEDGTIIITQALNAAKARLLEGGSLVKD
                     SPPHKVLKGHHQSVTSLLYPHGLSSKLDQSWMLSGDLDSCVILWDIFTEEILHKFPLE
                     AGPVTSLLMSPEKFKLRGEQIICCVCGDHSVALLHLEGKSCLLHARKHLFPVRMIKWH
                     PVENFLIVGCADDSVVIWEIETGTLERHETGERARIILNCCDDPQLVKSVLPIASETL
                     KHKSIEQRSSSPYQLGPLPCPGLQVESSCKVTDAKFCPRPFNVLPVKTKWSNVGFHIL
                     LFPLENLVELLLPPTPLSDVDSSSSFYGGEVLRRAKSTVEKKTLTLRKSKTACGPLSAE
                     ALAKPITRSLAQGDNTIKFSEENDGIKRQKMKISKKMQPKPSRKVDASLIIDTAKLF
                     LSCLLPWGVDKDLDYLCIKHLNLKLQGPISLGISLNEDNFSLMLPGWDLCNSGMIKD
                     YSGVNLFSRKVLDLSDKYTATLPNQVGIPRGLBNNCDSLRESDTIVLLSRLFLVNKL
                     VNMPLELACRVGSSFRMESIHNKWGRAGNDILNMSSFYSCLRNGNESHVPEADLSLL
                     KLISCWRDQSVQVTEAIQAVLLAEVQOHMKSLGKIPVNSQPVSMAENGNCEMKQMLPK
                     LEWTEELELQCVRNTLFLQTPVSPVKHDSNSNSANFQDVEDMPDRCALEESESPGEPR
                     HHSWIAKVCPCKVS"
     polyA_signal   5867..5872
                     /gene="DKFZp686F07114"
     polyA_site   5882
                     /gene="DKFZp686F07114"
ORIGIN

Query Match   87.5%; Score 1417.4; DB 9; Length 5900;
Best Local Similarity   96.1%; Pred. No. 1.3e-262;
Matches 1556; Conservative   0; Mismatches 50; Indels 13; Gaps 11;

Qy   1 TTAATCTACATATCTTATGTATATATATAAATATTGCCTAATCCATAGAAAAAGGATATAAAG 60
            ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   4289 TTAATCTACATATCTTATGTATATATATAAATATTGCCTAATCCATAGAAAAAGGATATAAAG 4348

Qy   61 TATTAAATATGTGATATAGAGTAGACTATATCTATCCATCTATGTATCTAACTAGCATTA 120
            ||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   4349 TATTAAATATGTGATATAGAGTAGACTATATCTATCCATCTATGTATCTAACTAGCATTA 4408

Qy   121 GTCACTTCAATTGAAGAAACATATCTCTGAGCAGTAGGAGCCAGCAGCCTCAGTCCTATGGTG 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4409 GTCACTTCAATTGAAGAAACATATCTCTGAGCAGTAGGAGCCAGCAGCCTCAGTCCTATGGTG 4468

Qy   181 GATGCAGTGGACAGGAGGGGGAATTGAAAAAGAGAACTATATAATTGAAAAGGGAT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4469 GATGCAGTGGACAGGAGGGGGAAATTGAAAAAGAGAACTATATAATTGAAAAGGGAT 4528

Qy   241 ATAAAGCATTAAATATGATATAGAGCTATATCTATGTATGTATGTATCTAACAGAGAAGTTC 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4529 ATAAAGCATTAAATATGATATAGAGCTATATCTATGTATGTATGTATCTAACAGAGAAGTTC 4588

Qy   301 AAGTCACTTCAATTAAAGAAACATTTTGAGCATGGACATGGACCAGCCTCAGGTCTTATGCTGG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4589 AAGTCACTTCAATTAAAGAAACATTTTGAGCATGGACATGGACCAGCCTCAGGTCTTATGCTGG 4648

Qy   361 GATGCAGTAGACAGGAGATGGGGAATTAGAAAAGAGAACTGTGTAATTGAAATGACGTG 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4649 GATGCAGTAGACAGGAGATGGGGAATTAGAAAAGAGAACTGTGTAATTGAAATGACGTG 4708

Qy   421 GGCTGCACCCTTAAGGAACTTATAATTAATGATGATCTGAATAAACATACCAGGATAAAG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4709 GGCTGCACCCTTAAGGAACTTATAATTAATGATGATCTGAATAAACATACCAGGATAAAG 4768

Qy   481 ATGTCAAATGAGTGTGACTCCCTAAAGTAGATTAAAGTGTGCATTCTTGTTTCCTAAA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4769 ATGTCAAATGAGTGTGACTCCCTAAAGTAGATTAAAGTGTGCATTCTTGTTTCCTAAA 4828

Qy   541 ATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTA 600
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4829 ATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTA 4888

Qy   601 ATATGATAATGCATGGAAATTATATCCCTTGAAAACTGATTGATAAATATATTCCCC 660
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4889 ATATGATAAT-CATGGAAATTATATCCCTTGAAAACTGATTGATAAATATATTCCCC 4947

Qy   661 CTCCCTTAGAAACAGTCAAAAGCACCTCAAACAAGTTCAAAATAAAGGAAGGTAGCAA 720
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4948 CTCCCTTAGAAACAGTCAAAAGCACCTCAAACAAGTTCAAAATAAAGGAAGGTAGCAA 5007

Qy   721 GTTAGGCGATGGATTATATTTTCTTCGGCTTCGTGTATACCCATTGGCCCAGGGCCTTATA 780
            |||||||||||||||||||||||||||||| |||||| ||||||||||||||||||
Db   5008 GTTAGGCGATGGATTATATTTTCTT--GCTTGTTGTATA--CCAGTTGTTGTCAAGGACATTATA 5065

Qy   781 AGGACTCCCAAAAGCATTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCTCAAAT 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5066 AGGACTCCCAAAAGCATTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCTCAAAT 5125

Qy   841 GAGGCATTTTAATTGTTACAATCTATTGGACGCTCAGGTTATGATATGTTTATGAAAA 900
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5126 GAGGCATTTTAATTGTTTAAATCTATTTGGACGCGTCAGGTTATGATATGTTTATGAAAA 5185

Qy   901 ATAAGCTTCATTATTCTTATAGCTACATTTATTCCCTTTAGAAACAAGAATAAC 960
            ||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   5186 ATAAGCTTCATTATTT-TTATAGCTACATCCTATATTCCCTTTAGAAACAAGAATAAC 5244

Qy   961 AATAAGTTTAATAGTTGCCATACTTAGCATTTATCAGGTCTCAATGAAACCATATTGAA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5245 AATAAGTTTAATAGTTGCCATACTTAGCATTTATCAGGTCTCAATGAAACCAATATTGAA 5304

Qy   1021 TCTTCGATAAATATTTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTG 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5305 TCTTCGATAAATATTTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTG 5364

Qy   1081 ACATTACTAAGTCGAAATGTTAGGATTTTCCGGCATCGCATGTTAGAATCTCTAAAATTT 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5365 ACATTACTAAGTCGAAATGTTAGGATTTTCCGGCATCGCATGTTAGAATCTCTAAAATTT 5424

Qy   1141 AAACATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATATGAATTCTGAAGGCCAATA 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5425 AAACATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATATGAATTCTGAAGGCCAATA 5484

Qy   1201 TCATACCATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTAGAGATATTCAAGCA 1260
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5485 TCATACCATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTAGAGATATTCAAGCA 5544

```
Qy   1261  ATGCTCTCCTAATATCCATACGCAAGTGTGTTTATGACACAAATTCACTAGTCTCTGGTTTA  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   5545  ATGCTCTCCTAATATCCATACGCAAGTGTGTTTATGACACAAATTCACTAGTCTCT-GTTTA  5603

Qy   1321  AAAATGAAATCTTTATATTGACTGGGGTCCCACATATTCCAGTAATTTCGTTATGAGA  1380
           ||||||||||||||||| |||||||  ||||||||||||| |||||||||||||||||
Db   5604  AAAATGAATTCTTTATATTGACT-GGTGTTCCACATATTTCAGTAATTTCGTTATGAGA  5662

Qy   1381  GGACTTGAAATAGCAAATTGCCCACACAGTAACTGGATAGATCACGTACGTGGTGATCA  1440
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   5663  GGACTTGAAATAGCAAATTG-CCACACAGTTAACTGATAGACCACGTACGTGGTGATCA  5721

Qy   1441  TAACCACTTGGTACTACACCCAGAAACTCAAAATTGTCTTTCTCCTGATGAGATATGGG  1500
           |||||||||||||||||||||||||||||||||||||||| |||||||||||  |||||
Db   5722  TAACCACTTGGTACTACACCCAGAAACTCAAAATTGTC-TTTCTCCTGATGAGATAT-GG  5779

Qy   1501  GTGTCCTTTGTACGTCTAGGGGCCTAGGGTACCCAAGTGAAGTGAATATATAAGCAAAAT  1560
           |||||||||||||||||| |||||||||   |||||||||||||||||| ||| ||||||
Db   5780  GTGCCCTTTGTACGTCTAGGCC---TAGGTAACCAGTGGAGTGATTATATTAGC-AAAT  5835

Qy   1561  GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAATTTAAAA  1619
           ||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||||
Db   5836  GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAAAAAAAAA  5894


RESULT 2
AC024061/c
LOCUS       AC024061              157042 bp    DNA     linear   PRI 10-JUL-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-232J12 map 15q21.3, complete
            sequence.
ACCESSION   AC024061
VERSION     AC024061.7  GI:14647263
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157042)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Sequencing of human chromosome 15 D15S146-D15S117 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 157042)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2000) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 157042)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2001) Multimegabase Sequencing Center, Institute
            for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
            98105, USA
COMMENT     On Jul 10, 2001 this sequence version replaced gi:13162487.
            ------------- Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@systemsbiology.org
            -------------  Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399
                Note:data AC066614 [Drafting center UWMSC] and AC066611
            [Drafting center UWMSC] were added for finishing.
FEATURES             Location/Qualifiers
```

```
source          1..157042
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="15"
                /map="15q21.3"
                /clone="RP11-232J12"
                /clone_lib="RPCI Human BAC library 11"
misc_feature    1..26818
                /note="overlap with RP11-80G8 AC066614 and
                RP11-348P16 AC066611. Data from overlapping BACs were
                combined and the consensus sequence determined from
                RP11-232J12 to the extent possible."
misc_feature    1..26818
                /note="overlap with RP11-80G8 AC066614"
unsure          8723
                /note="low quality data."
unsure          31605..31610
                /note="low quality data."
unsure          34826
                /note="low quality data."
unsure          42795..42800
                /note="low quality data."
unsure          63513..63514
                /note="low quality data."
unsure          124490..124492
                /note="low quality data."
misc_feature    143570..157042
                /note="overlap with RP11-348P16 AC066611."
ORIGIN

Query Match             87.4%;  Score 1415;  DB 9;  Length 157042;
Best Local Similarity   96.2%;  Pred. No. 2e-262;
Matches 1553;           Conservative 0;  Mismatches 49;  Indels 13;  Gaps 11;
```

```
Qy      1  TTAATCTACATATCTTATGTATATAAAATATTGCCTAATCCATAGAAAAAAAGGATATAAAG  60
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  17095  TTAATCTACATATCTTATGTATATAAAATATTGCCTAATCCATAGAAAAGAAAAGGATATAAAG  17036

Qy     61  TATTAAATATGTGATATATAGCNNNNNNNNNNNNNNNNNNNATAGGGAAGTTCAA  120
           |||||||||||||||||||||| ||||||||        |||||||||||||||
Db  17035  TATTAAATATGTGATATATAGCTATCATCATCTATGTATCTAATAGGGAAGTTCAA  16976

Qy    121  GTCACTCAATTGAAGAAACATATCTCTGAGCATAGGAGCAGCCTCCAGTTCCTATGGTGG  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16975  GTCACTCAATTGAAGAAACATATCTCTGAGCATAGGAGCAGCCTCCAGTTCCTATGGTGG  16915

Qy    181  GATGCAGTGGACAGGAGGGGGAAATTAGAAAAGAACTATATAATTGAAAAGGGAT  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16915  GATGCAGTGGACAGGAGGGGGAAATTAGAAAAGAACTATATAATTGAAAAGAAGGGAT  16856

Qy    241  ATAAAGCATTAAATATAATGATATAGCTATATCTGATGTATCTCAACAGAGAAGTTC  300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16855  ATAAAGCATTAAATATAATGATATAGCTATATCTGATGTATCTCAACAGAGAAGTTC  16796

Qy    301  AAGTCACTCAATTAAAGAAACACATTTTGAGCATGGGACCAGCCTCAGGTCTTATGCTGG  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16795  AAGTCACTCAATTAAAGAAACACATTTTGAGCATGGGACCAGCCTCAGGTCTTATGCTGG  16736

Qy    361  GATGCAGTAGACAGGAGGATGGGGAAATTAGAAAGAGAACTGTGTAATTGAAATGACGTG  420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16735  GATGCAGTAGACAGGAGGATGGGGAAATTAGAAAAGAGAACTGTGTAATTGAAATGACGTG  16676

Qy    421  GGCTGCACCCTTAAGGAACTTAAGGACTTATAATTAATGATCTGAATAACATACCAGGATAAAG  480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16675  GGCTGCACCCTTAAGGAACTTAAGGACTTATAATTAATGATCTGAATAAACATACCAGGATAAAG  16616

Qy    481  ATGTCAAATGAGTGTGACTCCCTAAAGTAGATTAAAGTGCATTCTTGTTCTTCCTAAA  540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16615  ATGTCAAATGAGTGTGACTCCCTAAAGTAGATTAAAGTGCATTCTTGTTCTTCCTAAA  16556

Qy    541  ATATGATTTTACTGCTTGAATTACATTGAGTTGAAGTTTAGAAACTAACATAGCATTA  600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16555  ATATGATTTTACTGCTTGAATTACATTGAGTTGAAGTTTAGAAACTAACATAGCATTA  16496
```

DEFINITION   Homo sapiens cDNA clone IMAGE:4608833, partial cds.
ACCESSION   BC058026
VERSION   BC058026.1 GI:34785600
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2150)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
             Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
             Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
             Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
             Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
             Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
             Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
             Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
             Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
             McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
             Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
             Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
             Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
             Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
             Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
             Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
             Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
             Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
             human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2 (bases 1 to 2150)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-SEP-2003) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: CLONTECH
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: National Institutes of Health Intramural
             Sequencing Center (NISC),
             Gaithersburg, Maryland;
             Web site: http://www.nisc.nih.gov/
             Contact: nisc_mgc@nhgri.nih.gov
             Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
             Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
             Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
             Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
             Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
             McDowell,C., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
             Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
             Young,A., Zhang,L.-H. and Green,E.D.
             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAL Plate: 48 Row: d Column: 11
             This clone was selected for full length sequencing because it
             passed the following selection criteria: Hexamer frequency ORF
             analysis.
FEATURES   Location/Qualifiers
     source   1..2150
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4608833"
                     /tissue_type="kidney"
                     /clone_lib="NIH_MGC_75"

```
Qy    601 ATATGAATAATGCATGGAAAATTATTATCCTTTGAAAACTGATTGATAAATATTCCCC 660
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db  16495 ATATGAATAAT-CATGGAAAATTATTATCCTTTGAAAACTGATTGATAAATATTCCCC 16437

Qy    661 CTCCTTAGAAACAGTCAAAAGCCACTTCAAACAAGTTCAAAATAAAGGAAGGTAGCAA 720
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db  16436 CTCCTTAGAAACAGTCAAAAGCCCACTTCAAACAAGTTCAAAATAAAGGAAGGTAGCAA 16377

Qy    721 GTTAGGCGATGGATTATATTTCTTGGCTGTGTGTATACCCCATTATA 780
          ||||||||||||||||||||||| ||||||||||||| ||  ||||||||||
Db  16376 GTTAGGCGATGGATTATATTTCTT-GCTGTGTGTATA-CCAGTTGTCAAGGACATTATA 16319

Qy    781 AGGACTCCCAAAAGCATTTGAAGAAGGCAATCAAATAAGTGTATGTCCTCTCAAAT 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16318 AGGACTCCCAAAAGCATTTGAAGAAGGCAATCAAATAAGTGTATGTCCTCTCAAAT 16259

Qy    841 GAGGCATTTTAATTGTTACAATCTATTTGGACCGCTCAGTTATGATATGTTTATGAAAA 900
          ||||||||||||||||||  ||||||||||| |||||||||||||||||||||||||||
Db  16258 GAGGCATTTTAATTGTTAAATCTATTTGGACGCTCAGTTATGATATGTTTATGAAAA 16199

Qy    901 ATAAGCTTCATTATTCTTATAGCTACAATCCTATTATTCCCTTTTAGAAACAAGAATAAC 960
          |||||||||||||||  || ||| |||||||||||||||||||||||||||||||||||
Db  16198 ATAAGCTTCATTATTT-TAATAGTTACAATCCTATTATTCCCTTTTAGAAACAAGAATAAC 16140

Qy    961 AATAAGTTTAATAGTTGCCATACTTAGCATTTATCAGTGTCTAATGAAACCAATATTGAA 1020
          |||| ||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  16139 AATAGTTTTAATAGTTGCCATACTTAGCATTTATCAGGTCTAATGAAACCAATATTGAA 16080

Qy   1021 TCTCTGATAAATATTTTCTGATGTTACTAGCTACTATGGGAAATTAGAACTCGCACACCCTG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  16079 TCTCTGATAAATATTTTCTGATGTTACTAGCTACTATGGGAAATTAGAACTCGGCACACCCTG 16020

Qy   1081 ACATTACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGAATCTCTAAAATT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16019 ACATTACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGAATCTCTAAAATT 15960

Qy   1141 AAACATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATATGAATTCTGAAGGCCAATA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15959 AAACATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATATGAATTCTGAAGGCCAATA 15900

Qy   1201 TCATACCAATAACTATGAGAAGCTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCA 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15899 TCATACCAATAACTATGAGAAGCTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCA 15840

Qy   1261 ATGCTCTCCTAATATCCATACGCGAAGTGTGTTTATGACACAAATTCACTAGTCTGGTTA 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15839 ATGCTCTCCTAATATCCATACGCGAAGTGTGTTTATGACACAAATTCACTAGTCTGGTTA 15781

Qy   1321 AAAATGAAATCTTTATATTGACTGGGGTGTCCCACATATTCCAGTATAATTCTGTTATGAGA 1380
          ||||||||| |||||||||||||  |||||| |||||||| ||||||||||||||||||||
Db  15780 AAAATGAATTCTTTATATTGACT-GGTGTTCCACATATTCAGTATAATTCTGTTATGAGA 15722

Qy   1381 GGACTGAAATAGCAAATTGCCACACAGTTAACTGGATAGATCACGTGGTGATCA 1440
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  15721 GGACTGAAATAGCAAATTG-CCACACAGTTAACTGGATAGATCACGTGGTGATCA 15663

Qy   1441 TAACCACTTGGTACTACACCCAGAAACTCAAAATTGTCTTTTCTCCTGATGAGATATGGG 1500
          |||||||||||||||||||||||||||||||||||||| || |||||||||||||| ||
Db  15662 TAACCACTTGGTACTACACCCAGAAACTCAAAATTGTC-TTCTCCTGATGAGATAT-GG 15605

Qy   1501 GTGTCCTTTGTACGTCTAGGGCCTAGGGCCCTAGCCTCAGGTAACCAGTGAATAATAAGCAAAT 1560
          |||||||||||||||||||||||   |||||||| |  |||||||||| |||||||| |||||
Db  15604 GTGTCCTTTGTACGTCTAGGGCC---TAGGTAACCAGTGAGTGATTATTAGC-AAAT 15549

Qy   1561 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAATTTATTTAAAAATTT 1615
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  15548 GTGTTTGTATCCGGAGTCTTCCTGTCATTGTAATAAAAATTTATTTAAAAATTT 15494
```

RESULT 3
BC058026
LOCUS   BC058026   2150 bp   mRNA   linear   PRI 16-SEP-2003

```
                /lab_host="DH10B"
                /note="Vector: pDNR-LIB"

ORIGIN

Query Match        87.4%;   Score 1414.2;  DB 9;  Length 2150;
Best Local Similarity 96.0%;  Pred. No. 6.6e-262;
Matches 1554;    Conservative   0;   Mismatches  52;   Indels  13;   Gaps  11;

Qy      1  TTAATTACATATCTTATGTATATAAATATTGCCTAATCCATAGAAAAAAGGATATAAAG   60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    522  TTAATTACATATCTTATGTATATAAATATTGCCTAATCCATAGAAAAAAGGATATAAAG  581

Qy     61  TATTAAAATATGTGATATATAGCNNNNNNNNNNNNNNNNNNNNNATAGGGAAGTTCAA  120
              ||||||||||||||||||||||||
Db    582  TATTAAAATATGTGATATATAGCTATATATCTATCTATGTCTAATATAGGGAAGTTCAA  641

Qy    121  GTCACTTCAATTGAAGAAACATATCTCTGAGCATAGAGCAGCCTCAGTTCCTATGGTGG  180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    642  GTCACTTCAATTGAAGAAACATATCTCTGAGCATAGAGCAGCCTCAGTTCCTATGGTGG  701

Qy    181  GATGCAGTGGACAGGAGGAGAGGGGGAAATTAGAAATAGAAGAACATATATAATTGAAA  240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    702  GATGCAGTGGACAGGAGGAGAGGGGGAAATTAGAAATAGAAGAACATATATAATTGAAA  761

Qy    241  ATAAAGCATTAAAATGAAATAAATATGATATCTATGTATGTATCTAACAGAGAAGTTC  300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    762  ATAAAGCATTAAAATGAAATAAATATGATATCTATGTATGTATCTAACAGAGAAGTTC  821

Qy    301  AAGTCACTTCAATTCAAATGGGAACAGCCAGCCTCAGTCTTATGCTGG  360
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db    822  AAGTCACTTCAATTCAAATGGGAACAGCCAGCCTCAGTCTTATGCTGG  881

Qy    361  GATGCAGTAGACAGGAGATGGGGAAATTAGAAAAAGAACTGTGTAATTGAAATGACGTG  420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    882  GATGCAGTAGACAGGAGATGGGGAAATTAGAAAAAGAACTGTGTAATTGAAATGACGTG  941

Qy    421  GGCTGCACCCTAAGGAACTATAATAATTAATGATCTGAATAAACATACCAGGATAAAG  480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    942  GGCTGCACCCTAAGGAACTATAATAATTAATGATCTGAATAAACATACCAGGATAAAG  1001

Qy    481  ATGTCAAATGAGTGTGACTCCCTAAAGTAAGTAAAGTGTGCATTCTTGTTCCTAAA  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1002  ATGTCAAATGAGTGTGACTCCCTAAAGTAAGTAAAGTGTGCATTCTTGTTCCTAAA  1061

Qy    541  ATATGATTTTACTGCTTGAAAATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTA  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1062  ATATGATTTTACTGCTTGAAAATTACATTTGAGTTGAAGTTTAGAAACTAACATAGTATTA  1121

Qy    601  ATATGAATAATGCATGGAAAATTATTATCCTTTGAAAACTGATTGATAAATAATATTCCCC  660
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   1122  ATATGAATAAT-CATGGAAAATTATTATCCTTTGAAAACTGATTGATAAATAATATTCCCC  1180

Qy    661  CTCCTTTGAAAACGTCAAAAGCCACTTCAAAACAAGTTTCAAAATAAAGGAAGGTAGCAA  720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1181  CTCCTTTAGAAAACAGTCAAAAGCCACTTCAAAACAAGTTTCAAAATAAAGGAAGGTAGCAA  1240

Qy    721  GTTAGGCGGATGGATTATATTTCTTGGCTGTTGTATACCCATTGGCCAGGGCCTTTATA  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1241  GTTAGGCGGATGGATGGCAATATTTCTTGGCTGTTGTATA-CCAGTTGTCAAGGACATTATA  1298

Qy    781  AGGACTCCCAAAAAGCATTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCTCAAAT  840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1299  AGGACTCCCAAAAAGCATTTTGAAGGATGGCAATATCAAATAAGTGTATGTCCTCTCAAAT  1358

Qy    841  GAGGCATTTTAATGTTCACAATCATCATTTGGACGGTCAGGTTATGATATGTTTATGAAAA  900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1359  GAGGCATTTTAATGTTTAATGTTAAAATCATCATTTGGACGGTCAGGTTATGATATGTTTATGAAAA  1418

Qy    901  ATAAGCTTCATTATTTCTTATAGCTACATCCTATTATTCCCTTTTAGAAAACAAGAATAAC  960
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1419  ATAAGCTTCATTATTT-TTATAGCTACATCCTATTATTCCCTTTTAGAAAACAAGAATAAC  1477

Qy    961  AATAAGTTTTAATAGTTGCCATACTTAGCATTATCAGGTCTAATGAAACCAATATTGAA  1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1478  AATAAGTTTTAATAGTTGCCATACTTAGCATTATCAGGTCTAATGAAACCAATATTGAA  1537

Qy   1021  TCTCTGATAAATATTTTCTGATGTTACTAGCTATGGGAAATTAGAACTGCACAACCCTG  1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1538  TCTCTGATAAATATTTTCTGATGTTACTAGCTATGGGAAATTAGAACTGCACAACCCTG  1597

Qy   1081  ACATTACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGAATCTCTAAAATTT  1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1598  ACATTACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGAATCTCTAAAATTT  1657

Qy   1141  AAACATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATAATGAATTCTGAAGGCCAATA  1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1658  AAACATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATAATGAATTCTGAAGGCCAATA  1717

Qy   1201  TCATACCATTAACTATGAAAGCTTTAATCCTAAAAATAGTTTTAGAGATATTCAAGCA  1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1718  TCATACCATTAACTATGAAAGCTTTAATCCTAAAAATAGTTTTAGAGATATTCAAGCA  1777

Qy   1261  ATGCTCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCTCGTTTA  1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1778  ATGCTCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCT-GTTTA  1836

Qy   1321  AAAATGAAATCTTTATATTGACTGGCGTGTCCCACATATTCCAGTAATTTCTGTTATGAGA  1380
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1837  AAAATGAATTCTTTATATTGACT-GGTGTTCCACATATTTCAGTAATTTCTGTTATGAGA  1895

Qy   1381  GGACTTGAAATAGCAAATTGCCCACACAGTTAACTGGATAGATCACGTACGTGGTGATCA  1440
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1896  GGACTTGAAATAGCAAATTG-CCACACAGTTAACTGGATAGACCACGTACGTGGTGATCA  1954

Qy   1441  TAACCACTTGGTACTACACCCAGAAACTCAAAATGTCTTTTCTCCTGATGAGATATGGG  1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1955  TAACCACTTGGTACTACACCCAGAAACTCAAAATGTC-TTTCTCCTGATGAGATAT-GG  2012

Qy   1501  GTGTCCTTTGTATCCGAGTCTTCCTGTCATTGTAATAAGTGAAGTGAATATATAAGCAAAT  1560
              |||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db   2013  GTGTCCTTTGTACGTCTAGGCC---TAGGTAACCAGTGGAGTGGATTATATTAGC-AAAT  2068

Qy   1561  GTGTTTGTATCCGAGTCTTCCTGTCATTGTCATTGTAATAAATTTATTTATTTAAAAATTTAAAA  1619
              |||||||||||||||||||||||||||||||||||||  |||
Db   2069  GTGTTTGTATCCGAGTCTTCCTGTCATTGTCATTGTAATAAAATTTATTCTAAAAAAAAAAA  2127


RESULT 4
HSM808902
LOCUS       HSM808902               5292 bp    mRNA    linear    PRI 30-AUG-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686L10117 (from clone DKFZp686L10117).
ACCESSION   BX648751
VERSION     BX648751.1  GI:34367915
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5292)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Deutschenbaur,S., Mewes,H.W.,
            Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German Human cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. This clone
            (DKFZp686L10117) is available at the RZPD in Berlin. Please contact
            the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/..
FEATURES             Location/Qualifiers
     source          1..5292
```

```
                                                                        Qy   841  GAGGCATTTTAATTGTACAATCATTATTGGACGCTCAGGTTATGATATGTTTATGAAAA    900
                                                                                  ||||||||||||||||| |||||| |||||||||||||||||||||||||||||||||||
                                                                        Db  4507  GAGGCATTTTAATTGTTAAAATCATTATTGGACGCTCAGGTTATGATATGTTTATGAAAA   4566

                                                                        Qy   901  ATAAGCTTCATTATTCTTATAGCTACATCCTATATTCCCTTTAGAAACAAGAGATAAC     960
                                                                                  ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
                                                                        Db  4567  ATAAGCTTCATTATTT-TTATAGCTACATCCTATATTCCCTTTAGAAACAAGAATAAC    4625

                                                                        Qy   961  AATAAGCTTAATAGTTGCCATACTTAGCGATTTATCAGGTCTAATGAAACCAATATTGAA  1020
                                                                                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                        Db  4626  AATAAGCTTAATAGTTGCCATACTTAGCGATTTATCAGGTCTAATGAAACCAATATTGAA  4685

                                                                        Qy  1021  TCTCTGATAAATATTTTCTGATGTTACTAGCTATGTGGAAATTAGAACTGGCACAACCCTG  1080
                                                                                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                        Db  4686  TCTCTGATAAATATTTTCTGATGTTACTAGCTATGTGGAAATTAGAACTGGCACAACCCTG  4745

                                                                        Qy  1081  ACATTACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCTAAAATTT  1140
                                                                                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                        Db  4746  ACATTACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCTAAAATTT  4805

                                                                        Qy  1141  AAACATTCCTGTTAAATGACTAAGGTTGCTTTATCAATATGAATTCTGAAGGCCAATA  1200
                                                                                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                        Db  4806  AAACATTCCTGTTAAATGACTAAGGTTGCTTTATCAATATGAATTCTGAAGGCCAATA  4865

                                                                        Qy  1201  TCATACCATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTAGAGATATTCAAGCA  1260
                                                                                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                        Db  4866  TCATACCATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTAGAGATATTCAAGCA  4925

                                                                        Qy  1261  ATGCTCTCCTAATATCCATAGCCAAGTGTGTTTATGACAAAATTCACTAGTCTGGTTTA  1320
                                                                                  |||||||||||||||||||||||||||||||||||||||||| ||||||||||| ||||
                                                                        Db  4926  ATGCTCTCCTAATATCCATAGCCAAGTGTGTTTATGACAAAATTCACTAGTCT-GTTTA  4984

                                                                        Qy  1321  AAAATGAAATCTTTATTATTGACTGGGTTGCCCACATATTCCAGTAGATCACGTGGTGATCA  1380
                                                                                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                        Db  4985  AAAATGAAATAGCAAATTG-CCACACAGTTAACTGACTAGATCACGTACGTGGTGATCA  5043

                                                                        Qy  1381  GGACTTGAAATAGCAAATTGACTGGGTTCCCACACAGTTAACTGGTCTTATGCTGTG  1440
                                                                                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                                        Db  5044  GGACTTGAAATAGCAAATTG-CCACACAGTTAACTGACTAGATCACGTACGTGGTGATCA  5102

                                                                        Qy  1441  TAACCACTTGGTACTACACCCAGAAACTCAAAATTGTCTTTTCCTGATGAGAGATATGGG  1500
                                                                                  |||||||||||||||||||||||||||||||||||||||||| | |||||||||||
                                                                        Db  5103  TAACCACTTGGTACTACACCCAGAAACTCAAAATTGTC-TTTCTCCTGATGAGAGATAT-GG  5160

                                                                        Qy  1501  GTGTCCTTTTGTACGTCTGAGGGCTCAGGGTACCCAAGTGAGTGAATATAAGCAAAAT  1560
                                                                                  ||||||||||||||||||||||||||||   |||||||||||||||||| |||||||
                                                                        Db  5161  GTGTCCTTTTGTACGTCTCAGGCC---TAGGTAACCAGTGGAGTGATTATATTAGC-AAAT  5216

                                                                        Qy  1561  GTGTTTGTATCCAGAGTCTTCCTGTCATTGATTAAAGTGTGCATTCTTGTTCTTGTTCCTAAA  1619
                                                                                  |||||||||||||||||||||||||||||||||| |||||||| |||| |||
                                                                        Db  5217  GTGTTTGTATCCAGAGTCTTCCTGTCATTGATTAAAGTGTGCATTCTTGTTTATTTAAAAAAAGA  5275
```

```
RESULT 5
HSM808720
LOCUS       HSM808720               6104 bp    mRNA     linear   PRI 30-AUG-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686M05115 (from clone DKFZp686M05115).
ACCESSION   BX648571
VERSION     BX648571.1  GI:34367733
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6104)
  AUTHORS   Bloecker,H., Boecher,M., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German Human cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
  COMMENT   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
```

```
                                        /organism="Homo sapiens"
                                        /mol_type="mRNA"
                                        /db_xref="taxon:9606"
                                        /map="15q21.3"
                                        /clone="DKFZp686L10117"
                                        /tissue_type="human fetal kidney"
                                        /clone_lib="686 (synonym: h1cc3). Vector pSport1_Sfi; host
                                        DH10B; Sites SfiIA + SfiIB"
                                        /dev_stage="fetal"
     polyA_signal       5248..5253
     polyA_site         5263
ORIGIN

Query Match     87.3%;  Score 1412.6;  DB 9;  Length 5292;
Best Local Similarity   95.9%;  Pred. No. 1.1e-261;
Matches 1553;  Conservative   0;  Mismatches   53;  Indels   13;  Gaps   11;
```

```
Qy     1  TTAAATCACATATGTGTATATATAATATTGCCTAATCCATAGAAAAAAGGATATAAAG     60
          |||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db  3670  TTAAATCACATATCTTATGTGTATATATAATATTGCCTAATCCATAGAAAAAAAGGGTATAAAG   3729

Qy    61  TATTAAAATGTGATATATAGCNNNNNNNNNNNNATAGGGAAGTTCAA     120
          |||||||||||||||||||||           ||||||||||||||||
Db  3730  TATTAAAATGTGATATATAGCTATATCTATCTATGTATCTAATAGGGAAGTTCAA     3789

Qy   121  GTCACTTCAATTGAAGAAACATATCTCTGAGCATAGGAGCAGCCTCAGGTCCTATGGTGG   180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3790  GTCACTTCAATTGAAGAAACATATCTCTGAGCATAGGAGCAGCCTCAGGTCCTATGGTGG   3849

Qy   181  GATGCAGTGGACAGGAGGGGGGGAAATTAGAAATGACTATATAATTGAAAAAGGGAT    240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3850  GATGCAGTGGACAGGAGGGGGGGAAATTAGAAGAGAACTATATAATTGAAAAAGGGAT   3909

Qy   241  ATAAAGCATTAAATATAGATAATGDATATCTTGTAGTCTGATCTAACAGAGAAGTTC    300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3910  ATAAAGCATTAAATATAGATAATGATATCTTGTAGTCTGATCTAACAGAGAAGTTC   3969

Qy   301  AAGTCACTTCAATTAAAGAAACATTTGAGCATGGGACCAGCCTCAGGTCTTATGCTGTG   360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3970  AAGTCACTTCAATTAAAGAAACATTTGAGCATGGGACCAGCCTCAGGTCTTATGCTGTG   4029

Qy   361  GATGCAGTAGACAGGAGGATGGGGAAATTAGAAGAGAACTGTGTAAATGACGTG     420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4030  GATGCAGTAGACAGGAGGATGGGGAAATTAGAAGAGAACTGTGTAATTGAAACGTG    4089

Qy   421  GGCTGCACCCTTAAGGAACTTATATAATTAATTAATGATGAATCTGAATAAACATACAGGATAAAG   480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4090  GGCTGCACCCTTAAGGAACTTATATAATTAATTGATGATGAATCTGAATAAACATACCAGGATAAAG  4149

Qy   481  ATGTCAAATGAGTGTGACTCCCTTAAAGTAAGTGCATTCTTTGTTCCTAAA    540
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4150  ATGTCAAATGAGTGTGACTCCCTTAAAGTAAGTGCATTCTTTGTTCCTAAA   4209

Qy   541  ATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTA   600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4210  ATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTA  4269

Qy   601  ATATGAATAATGCATGGAAATTATATCCTTTGAAAACTGATTGATAAATATTCCCC    660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4270  ATATGAATAAT-CATGGAAATTATATCCTTTGAAAACTGATTGATAAATATTCCCC   4328

Qy   661  CTCCTTTAGAAACAGTCAAAAGCCACTTCAAACAAGTTCAAAATAAAGGAAGGTAGCAA   720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4329  CTCCTTTAGAAACAGTCAAAAGCCCACTTCAAACAAGTTCAAAATAAAGGAAGGTAGCAA  4388

Qy   721  GTTAGGCGATGGATTATATATTTCTTGGCTTGTGTTGTATACCCATTGGCCAGGCCTTTATA   780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4389  GTTAGGCCATGGATTATATATTTCTT-GCTTGTTGTATA-CCAGTTGTCAAGGACATTATA  4446

Qy   781  AGGACTCCCAAAAGCATTTGAAGAATGGCAATATCAAATAAGTCATGTCCTCTCAAAT   840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4447  AGGACTCCCAAAAGCATTTGAAGAATGGCAATATCAAATAAGTCATGTCCTCTCAAAT  4506
```

```
          sequenced by GBF (National Research Centre for Biotechnology Ltd.,
          Braunschweig/Germany) within the cDNA sequencing consortium of the
          German Genome Project.
          This clone (DKFZp686M05115) is available at the RZPD in Berlin.
          Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
          Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
          information about the clone and the sequencing project is available
          at http://mips.gsf.de/proj/cDNA/.
FEATURES        Location/Qualifiers
     source     1..6104
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /map="15q21.3"
                /clone="DKFZp686M05115"
                /tissue_type="human fetal kidney"
                /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                DH10B; sites SfiIA + SfiIB"
                /dev_stage="fetal"
     polyA_signal   6067..6072
     polyA_site     6082
ORIGIN

Query Match          87.3%;    Score 1412.6;   DB 9;    Length 6104;
Best Local Similarity 95.9%;   Pred. No. 1/e-261;
Matches 1553;  Conservative   0;   Mismatches  53;   Indels  13;   Gaps  11;

Qy       1 TTAATCTACATATCTTATGTATATAAATATTGCCTAATCCATAGAAAAAAGAATAAAG    60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4489 TTAATCTACATATCTTATGTATATAAATATTGCCTAATCCATAGAAAAAAGAATAAAG    4548

Qy      61 TATTAAATATGTGATATATAGCNNNNNNNNNNNNNNNNNNNNNNNATGGGAAGTTCAA   120
           |||||||||||||||||||||||
Db    4549 TATTAAATATGTGATATATAGCTATATCTCATCTCATCATGTATCTAATGAGGAAGTTCAA   4608

Qy     121 GTCACTTCAATTGAAGAAACATATCTCTGAGCATAGGAGCAGCCTCAGGTCCTATGTGG   180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4609 GTCACTTCAATTGAAGAAACATATCTCTGAGCATAGGAGCAGCCTCAGGTCCTATGTGG   4668

Qy     181 GATGCAGTGGACAGGAGGGGGAAATTAGAAATAGAACTATATAATGAAAAGGGAT   240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4669 GATGCAGTGGACAGGAGGGGGAAATTAGAAATAGAACTATATAATGAAAAGGGAT   4728

Qy     241 ATAAAGCATTAAATAATATGATATATAGCTATATCTCATGTATGTATCAACAGAGAAGTTC   300
           |||||||||||||||||||||||||||||||||||||||||||||
Db    4729 ATAAAGCATTAAATAATATGATATATAGCTATATCTCATGTATGTATCAACAGAGAAGTTC   4788

Qy     301 AAGTCACTTCAATTAAAGAAACATTTTGAGCATGGACCAGCCTCAGGTCTTATGCTGG   360
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    4789 AAGTCACTTCAATTAAAGAAACATTTTGAGCATGGACCAGCCTCAGGTCTTATGCTGG   4848

Qy     361 GATGCAGTAGACAGGAGGATGGGGAAATGGGGAAATTAGAAAAGAGACTGTGTAATTGAAATGACGTG   420
           |||||||||||||||||||||||||||||||||||||||||
Db    4849 GATGCAGTAGACAGGAGGATGGGGAAATGGGGAAATTAGAAAAGAGACTGTGTAATTGAAATGACGTG   4908

Qy     421 GGCTGCACCCTTAAGGAACTTATAATTAATATAATTAATGATCTGAATAAACATACCAGGATAAAG   480
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    4909 GGCTGCACCCTTAAGGAACTTATAATTAATATAATTAATGATCTGAATAAACATACCAGGATAAAG   4968

Qy     481 ATGTCAAATGAGTGTGACTCCCTTAAAGTAGTGCATTCCTTTGTTCCTAAA   540
           |||||||||||||||||||||||||||||||||||||||||||
Db    4969 ATGTCAAATGAGTGTGACTCCCTTAAAGTAGTGCATTCCTTTGTTCCTAAA   5028

Qy     541 ATATGATTTTACTCGTCTTGAAATTACATTTGAGTTGAAGTTGAAACTAACATAGACATTA   600
           |||||||||||||||||||||||||||||||||||||||||||||||
Db    5029 ATATGATTTTACTCGTCTTGAAATTACATTTGAGTTGAAGTTGAAACTAACATAGTATTA   5088

Qy     601 ATATGAATAATGCATGGAAAATTATTATCCTTTGAAAACTGATTGATAAAATATTCCCC   660
           |||||||||||||||||||||||||||||||||||||||||||
Db    5089 ATATGAATAAT-CATGGAAAATTATTATCCTTTGAAAACTGATTGATAAATATATTCCCC   5147

Qy     661 CTCCTTTAGAAACAGTCAAAAGCCACTTCAAACAGTTCAAACAATAAAGGAAGGTAGCAA   720
           |||||||||||||||||||||||||||||||||||||||
```

```
Db    5148 CTCCTTTAGAAACAGTCAAAAGCCACTTCAAACAGTTCAAACAGTTCAAATAAAGAAGGTAGCAA   5207

Qy     721 GTTAGGCGATGGATTATATTTTCTGGCTGTGTTGTATACCCATTGGCCAGGGCCTTTATA   780
           ||||||||||||||||||||||||||| | |||||||||||||||
Db    5208 GTTAGGCGATGGATTATATTTTTCT-GCTTGTGTATA-CCAGTTGTCAAGGACATTATA   5265

Qy     781 AGGACTCCCAAAAGCATTTGAAGAATGGCAATAACAAATAAGTATGTGTCCTCTCCAAAT   840
           |||||||||||||||||||||||||||||||||||||||||||||
Db    5266 AGGACTCCCAAAAGCATTTGAAGGATGGCAATAATCAAATCAAATAAGTGTATGTCCTCTCCAAAT   5325

Qy     841 GAGGGCATTTTTAATTGTTACAATCTATTGGACGCTCAGGTTGATGATATGTTATGAAAA   900
           |||||||||||||||||||||||||||||||||||||||||||||
Db    5326 GAGGGCATTTTTAATTGTTATTGGACGGTCAGGTTATGATATGTTTATGAAAA   5385

Qy     901 ATAAGCTTCATTTATTCTTATAGCTACATCCTATATTCCCTTTAGAAACAAGAATAAC   960
           ||||||||||||||||||||||| ||||||||||||||||||||||
Db    5386 ATGAGCTTCATTATTT-TTATAGCTACATCCTATTATTCCCTTTAGAAACAAGAATAAC   5444

Qy     961 AATAAGTTTAATAGTTGCCATACTTAGCATTATTCAGGTCTAATGAAACCAATATTGAA   1020
           |||||||||||||||||||||||||||||||||||||||||||||
Db    5445 AATAAGTTTAATAGTTGCCATACTTAGCATTTATCAGGTCTAATGAAGCCAATATTGAA   5504

Qy    1021 TCTCTGATAAATATTTTCTGATGTTACTAGCTAGCTAGTATGGGAAATTAGAACTGGCACAACCCTG   1080
           ||||||||||||||||||||||||||||||||||||||||||
Db    5505 TCTCTGATAAATATTTTCTGATGTTACTAGCTAGCTAGTATGGGAAATTAGAACTGGCACAACCCTG   5564

Qy    1081 ACATTACTAAGTGGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTTCTAAAATTT   1140
           ||||||||||||||||||||||||||||||||||||||||||
Db    5565 ACATTACTAAGTGGGAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTTCTAAAATTT   5624

Qy    1141 AAACATTCCTGTTAAATGACTAAGGTTGCTTTATCAATATGAATTCTGAGGCCAATA   1200
           ||||||||||||||||||||||||||||||||||||||||
Db    5625 AAACATTCCTGTTAAATGACTAAGGTTGCTTTATCAATATGAATTCTGAGGCCAATA   5684

Qy    1201 TCATACCATTAACTATGAAAGCTTTAATTCCTAAAAAATAGTTTAGAGATATTCAAGCA   1260
           |||||||||||||||||||||||||||||||||||||||||||
Db    5685 TCATACCATTAACTATGAAAGCTTTAATTCCTAAAAAATAGTTTAGAGATATTCAAGCA   5744

Qy    1261 ATGCTCTCCAATATCCATACGCAAGTGTGTTTATGACAAAATTCACTAGTCTGGTTTA   1320
           |||||||||||||||||||||||||||||||||||||||||
Db    5745 ATGCTCTCCAATATCCATACGCAAGTGTGTTTATGACAAAATTCACTAGTCT-GTTTA   5803

Qy    1321 AAAATGAAATCTTTATTGACTGGGGTCCCACATATTCCAGTAATTCTGTTATGAGA   1380
           ||||||||||||||||||||||||||||||||||||||||||
Db    5804 AAAATGAATTCTTTATTGACT-GGTGTTCCACATATTCCAGTAATTTCGTTATGAGA   5862

Qy    1381 GGACTTGAAATAGCAAATTGCCCACACAGTTAACTGGTGTGATCACGTACGTGGTGATCA   1440
           ||||||||||||||||||||||||||||||||||||||||||||
Db    5863 GGACTTGAAATAGCAAATTG-CCACACAGTTAACTGGATAGACCACGTACGTGGTGATCA   5921

Qy    1441 TAACCACTTGGTACTACACCCAGAACTCAAAAATGTCTTTTCTCCTGATGAGATATGGG   1500
           |||||||||||||||||||||||||||||||||||||||||||||
Db    5922 TAACCACTTGGTACTACACCCAGAACTCAAAAATGTC-TTTCCTCGATGAGATGAGATAT-GG   5979

Qy    1501 GTGTCCTTTTGTACGTTGTACGTTGTACGTTCTAGGGCTCAGGGCC---TAGGTAACCAGTGGATTGTAA   1560
           ||||||||||||||||||||||||||||||||||||||||||||
Db    5980 GTGTCCTTTTGTACGTTGTACGTTGTACGTTCTAGGCC---TAGGTAACCAGTGGATTGATTATTAGC-AAAT   6035

Qy    1561 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAATTTATTTATTTAAAA   1619
           |||||||||||||||||||||||||||||||||||||||||
Db    6036 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAATTTATTTAAAAAAAAA   6094
```

```
RESULT 6
AC066614/c
LOCUS       AC066614              176626 bp    DNA      linear    PRI 09-MAR-2002
DEFINITION  Homo sapiens chromosome 15 clone RP11-80G8 map 15q21.3, complete
            sequence.
ACCESSION   AC066614
VERSION     AC066614.8  GI:19310269
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 176626)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE   Sequencing of human chromosome 15 D15S146-D15S117 region
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 176626)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P. Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE   Direct Submission
  JOURNAL   Submitted (25-APR-2000) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3 (bases 1 to 176626)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-MAR-2002) Multimegabase Sequencing Center, Institute
            for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
            98105, USA
COMMENT   On Mar 9, 2002 this sequence version replaced gi:13702797.
            ------------- Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@systemsbiology.org
            ------------- Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399
                Note:data from overlapping clones AC087832 [drafting centers
                WIBR,UMSC],AC026598 [drafting center WIBR],and AC024061 drafting
                center UMWSC] were added for finishing.
FEATURES   Location/Qualifiers
     source   1..176626
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q21.3"
                     /clone="RP11-80G8"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature   1..141109
                     /note="This BAC overlaps RP11-6417 AC087832,RP11-418B13
                     AC026598, and RP11-232Q12 AC024061.Data from overlapping
                     BACs were combined and the consensus sequence determined
                     from RP11-80G8 to the extent possible."
     misc_feature   1..141109
                     /note="overlap with RP11-418B13 AC026598."
     misc_feature   1..49308
                     /note="overlap with RP11-6417 AC087832."
     misc_feature   149918..176626
                     /note="overlap with RP11-232Q12 AC024061."
     misc_feature   156003
                     /note="removed TN10 insertion from some clones of
                     RP11-80G8."
ORIGIN

Query Match   87.2%; Score 1411.8; DB 9; Length 176626;
Best Local Similarity 96.0%; Pred. No. 8.3e-262;
Matches 1551; Conservative 0; Mismatches 51; Indels 13; Gaps 11;

Qy   1 TTAATCTACAATATCTTATGTGATATAAGCTATATCTATCTATCCATTAGAAAAAGGATATAAAG   60
             |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   166906 TTAATCTACAATATCTTATGTGATATAAGCTATATCTATCTATCCATTAGAAAAAGGATATAAAG 166847

Qy   61 TATTAAAATATGTGATATAAGCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATAGGGAAGTTCAA 120
             |||||||||||||||||||||||
Db   166846 TATTAAAATATGTGATATAAGCTATATCTATCTATCCATTAGTAATCTCAATAGGGAAGTTCCAA 166787

```
            /clone_lib="NT2RI2"
            /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
            precursor cells treated 2-weeks mitotic inhibitor after
            5-weeks retinoic acid (RA) induction.~majorly NT2 neuron"

ORIGIN

Query Match        87.1%;    Score 1410.2;  DB 9;  Length 2455;
Best Local Similarity   96.3%;   Pred. No. 3.8e-261;
Matches 1547;   Conservative   0;   Mismatches 47;   Indels 13;   Gaps 11;

Qy     1 TTAATCTACATATCTTATGTATATAAATAGTTTGCCAATCCATAGAAAAAAGAAGATATAAAG 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   862 TTAATCTACATATCTTATGTATATAAATAGTTTGCCAATCCATAGAAAAAAGAAGATATAAAG 921

Qy    61 TATTAAATATGTGATATATAGAGCNNNNNNNNNNNNNNNNNNNATAGGGAAGTTCAA 120
         |||||||||||||||||||||||
Db   922 TATTAAATATGTGATATATAGAGCTATATATCATCTATATGTCTAATAGGGAAGTTCAA 981

Qy   121 GTCACTTCAATTGAAGAAACATATCCTGAGCATAGGAGGAGCCTCCAGTCCTCATGGTGG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   982 GTCACTTCAATTGAAGAAACATATCCTGAGCATAGGAGGAGCCTCCAGTCCTCATGGTGG 1041

Qy   181 GATGCAGTGGACAGGAGGAGGGGGAAATTAGAAAGAAGAACTATATAATTGAAAAAGGGAT 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1042 GATGCAGTGGACAGGAGGAGGGGGAAATTAGAAAGAAGAACTATATAATTGAAAAAGGGAT 1101

Qy   241 ATAAAGCATAAATATAATATGATATAGCTATATCTATATCTATGTATGTATGTATCTAACAGAGAAGTTC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1102 ATAAAGCATAAATATAATATGATATAGCTATATCTATATCTATGTATGTATGTATCTAACAGAGAAGTTC 1161

Qy   301 AAGTCACTTCAATTAAAGAAACATTTTGAGCATGGACCAGCCTCAGTTCTTATGCTGTG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1162 AAGTCACTTCAATTAAAGAAACATTTTGAGCATGGACCAGCCTCAGTTCTTATGCTGTG 1221

Qy   361 GATGCAGTAGACAGGAGATGGGGAAATTAGAAATTAGAAAAGAACTGTGTAATTGAAATGACGTG 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1222 GATGCAGTAGACAGGAGATGGGGAAATTAGAAATTAGAAAAGAACTGTGTAATTGAAATGACGTG 1281

Qy   421 GGCTGCACCCTTAAGGAACTATAATTAATGATCTGAATAAACATACCAGGATAAAG 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1282 GGCTGCACCCTTAAGGAACTATAATTAATGATCTGAATAAACATACCAGGATAAAG 1341

Qy   481 ATGTCAAATGAGTGGACTCCCTTAAAGTAGTAATTAAAGTGTGCATTCTTGTTCCTAAA 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1342 ATGTCAAATGAGTGGACTCCCTTAAAGTAGTAATTAAAGTGTGCATTCTTGTTCCTAAA 1401

Qy   541 ATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTA 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1402 ATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAGTTTAGAAACTAACATAGCATTA 1461

Qy   601 ATATGAATAAATGCATGGAGAAAATTATTATCCTTGAAAACTGATTGATTAAATATATTCCCC 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1462 ATATGAATAAAT-CATGGAGAAAATTATTATCCTTGAAAACTGATTGATTAAATATATTCCCC 1520

Qy   661 CTCCCTTAGAAACAGTCAAAAGCCACTTCAAACAAGTTCAAAATAAAGGAAGGTAGCAA 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1521 CTCCCTTAGAAACAGTCAAAAGCCACTTCAAACAAGTTCAAAATAAAGGAAGGTAGCAA 1580

Qy   721 GTTAGGCGATGGATTATATTTCTTGGCTTGTGTATACCCATTGGCCAGGCGCCTTTATA 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1581 GTTAGGCGATGGATTATATTTCTT-GCTTGTGTATA-CCAGTTGTGTCAAGGACATTATA 1638

Qy   781 AGGACTCCCAAAAGCATTTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCTCAAAT 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1639 AGGACTCCCAAAAGCATTTTGAAGGATGGCAATATCAAATAAGTGTATGTCCTCTCAAAT 1698

Qy   841 GAGGCATTTTAATTGTTACAATCTATTTGGACGCTCAGGTTATGATATGTTTATGAAAA 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1699 GAGGCATTTTAATTGTTAAAATCTATTTGGACGCTTCAGGTTATGATATGTTTATGAAAA 1758

Qy   901 ATAAGCTTCATTATTCTTATAGCTACATCCTTATTATTCCCTTTTGAAAACAAGAATAAC 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1759 ATAAGCTTCATTATTT-TTATAGCTACATCCTTATTATTCCCTTTTGAAAACAAGAATAAC 1817
```

```
Qy  1201 TCATACCATTAACTACGAAAGCTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCA 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165710 TCATACCATTAACTACGAAAGCTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCA 165651

Qy  1261 ATGCTCTCCTAATATCCATACGGCAAGTGTGTTTATGACACAAATTCACTACGTCTGTTTA 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165650 ATGCTCTCCTAATATCCATACGGCAAGTGTGTTTATGACACAAATTCACTACGTCT-GTTTA 165592

Qy  1321 AAAATGAAATCTTTATATTGACTGGGGTGTCCCACATATTCCAGTAATTCTGTTATGAGA 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165591 AAAATGAATTCTTTATATTGACT-GGTGTCCCACATATTCAGTAATTCTGTTATGAGA 165533

Qy  1381 GGACTTGAAATAGCAAATTGCCCACACAGTTAACTGGATAGATCACGTACGTGGTGATCA 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165532 GGACTTGAAATAGCAAATTG-CCACACAGTTAACTGGATAGAGCACGTACGTGGTGATCA 165474

Qy  1441 TAACCACTTGGTACTACACCCAGAAAACTCAAAATGTCTTTTCTCCTGATGAGATATGGG 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165473 TAACCACTTGGTACTACACCCAGAAAACTCAAAATTGTC-TTTCTCCTGATGAGATAT-GG 165416

Qy  1501 GTGTCCTTTGTACGTCCAGGGCCTAGGGTACCCAAGTGAAGTGAATAATATAAGCAAAAT 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165415 GTGTCCTTTGTACGTCCAGGCC---TAGGTAACCAGTGGAGTGAATATATAAGC-AAAT 165360

Qy  1561 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAATTATTGTAATAAAAATTT 1615
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 165359 GTGTTTGTATCCGGAGTCTTCCTGTCATTGTAATAAAAATTATTTAATAAAAATTT 165305
```

```
RESULT 7
AK123805
LOCUS       AK123805                2455 bp    mRNA    linear    PRI 09-SEP-2003
DEFINITION  Homo sapiens cDNA FLJ41811 fis, clone NT2RI2010283.
ACCESSION   AK123805
VERSION     AK123805.1  GI:34529435
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
            Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakanatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,K., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE       NEDO human cDNA sequencing project
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 2455)
AUTHORS     Isogai,T. and Yamamoto,J.
TITLE       Direct Submission
JOURNAL     Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..2455
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RI2010283"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
```

COMMENT   Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
            NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.

FEATURES   Location/Qualifiers
     source   1..3493
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CTONG3007528"
                     /tissue_type="tongue, tumor tissue"
                     /clone_lib="CTONG3"
                     /note="cloning vector: pME18SFL3"
     CDS   177..962
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC86058.1"
                     /db_xref="GI:34531120"
                     /translation="MLPGWDLCNSGMIKDYSGVNLFSRKVLDLSDKYTATLPNQVGIP
                     RGLENNCDSLRESDTIVYLLSRLFIVNKLVNMPLELACRVGSSPMESIHNKMRGAGN
                     DILNMSSFYSCLRNGKNESHVPEADLSLLKLISCWRDQSVQVTEAIQAVLLAEVQQHM
                     KSLGKIPVNSQPVSMAENGNCEMKQMLPKLEWTELELELQCVRNTLPLQTPVSPVKHDS
                     NSNSANFQDVEDMPDRCALEESESPGEPRHHSWIAKVCPCKVS"

ORIGIN
Query Match   87.0%; Score 1408.6; DB 9; Length 3493;
Best Local Similarity 96.2%; Pred. No. 7.2e-261;
Matches 1546; Conservative 0; Mismatches 48; Indels 13; Gaps 11;

Qy   1 TTAATCACATATCTATGTATATAAATATTGCCTAATCCATAGAAAAAAGATATAAAG   60
          |||||||||||| ||||||||||||||||||||||||||||||||||||| ||||||||
Db   1900 TTAATCACATATCTTATGTATATAAATATTGCCTAATCCATAGAAAAAAGGATATAAAG   1959

Qy   61 TATTAAATATGTGATATAGCNNNNNNNNNNNNNNNNNNNATAGGGAAGTTCAA   120
          |||||||||||||||||||| ||||||||||||||
Db   1960 TATTAAATATGTGATATAGCTATATCTATCTATCTATGTATCTAATAGGGAAGTTCAA   2019

Qy   121 GTCACTTCAATTGAAGAAACAATATCTCTGAGCATAGGAGCAGCCAGGTCCATGGTGG   180
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   2020 GTCACTTCAATTGAAGAAACATATCTCTGAGCATAGGAGCAGCCAGGTCCATGGTGG   2079

Qy   181 GATGCAGTGGACAGGAGGAGGGGGAAATTAGAAAAAGAGAACTATATAATTGAAAAAGGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   2080 GATGCAGTGGACAGGAGGAGGGGGAAATTAGAAAAAGAGAACTATATATATTGAAAAAGGGAT 2139

Qy   241 AATAAAGCATTAAATATGATATAGCTATATCTATGTATGTATCTAACAGGAAGTTC   300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2140 AATAAAGCATTAAATATGATATAGCTATATCTATGTATGTATCTAACAGGAAGTTC   2199

Qy   301 AAGTCACTTCAATTAAAGAAAACATTTTTGAGCATGGGACCAGCCTCTATGCTGG   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2200 AAGTCACTTCAATTAAAGAAAACATTTTTGAGCATGGGACCAGCCTCTATGCTGG   2259

Qy   361 GATGCAGTAGACAGGAGGATGGGGAAATTAGAAAAAGAGAACTGTGTAATTGAAATGACGTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2260 GATGCAGTAGACAGGAGGATGGGGAAATTAGAAAAAGAGAACTGTGTAATTGAAATGACGTG 2319

Qy   421 GGCTGCACCCCTTAAGGACGTTAAATAATGGAATCTGAATAAACATACCAGGATAAAG   480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2320 GGCTGCACCCCTTAAGGACGTTAAATAATGGAATCTGAATAAACATACCAGGATAAAG   2379

Qy   481 ATGTCAAATGAGTGTGACTCCCTAAAGTAAGTAGGATTAAAGTGTGCATTCTTGTTCCTAAA 540
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   2380 ATGTCAAATGAGTGTGACTCCCTAAAGTAAGTAGAGATTAAAGTGTGCATTCTTGTTCCTAAA 2439

Qy   541 ATATGATTTTACTGCTTGAAATTACATTTGAGTTGAAGTTAGAAACTAACATAGCATTA   600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   961 AATAAGTTTTAAATAGTTGCCATACTTAGCCATTATCAGGTCTAATGAAACCAATATTGAA   1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1818 AATAAGTTTTAAATAGTTGCCATACTTAGCCATTATCAGGTCTAATGAAACCAATATTGAA   1877

Qy   1021 TCTCTGATAAATATTTTCTGATGTTGACTAGCTATGGGAAATTAGGAACTGGCACACCCTG 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1878 TCTCTGATAAATATTTTCTGATGTTGACTAGCTATGGGAAATTAGGAACTGGCACACCCTG   1937

Qy   1081 ACATTACTAGTGGAAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCTAAAATT   1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1938 ACATTACTAGTGGAAAATGTTAGGATTTTCGGCATCGCATGTTAGAATCTCTAAAATT   1997

Qy   1141 AAACATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATGAATTCTGAAGGCCAATA   1200
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   1998 AAACATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATCAATATGAATTCTGAAGGCCAATA 2057

Qy   1201 TCATACCATTAACTATGAAAGCTTTAATTCCTAAAAATAGTTTAGAGATATTCAAGCA   1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2058 TCATACCATTAACTATGAAAGCTTTAATTCCTAAAAATAGTTTAGAGATATTCAAGCA   2117

Qy   1261 ATGCTCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCTGGTTTA   1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   2118 ATGCTCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCT-GTTTA   2176

Qy   1321 AAAATGAAATCTTTATATTGACTGGGTGTGCCCACATATTCCAGTGAATTCTGTTATGAGA 1380
          |||||||||||||||||||||| | |||| ||||||||| ||| |||| ||||||||||||
Db   2177 AAAATGAAATCTTTATATTGACT-GGTGTTCCACATATTTCAGTTATTTCTGTTATGAGA   2235

Qy   1381 GGACTTGAAATAGCAAAATTGCCCACACAGTTAACTGGATAGATCACGTACGTGTGATCA   1440
          |||||||||||||||||||| |||||||||||||||||||||| ||||||||||||||||
Db   2236 GGACTTGAAATAGCAAAATTG-CCACACAGTTAACTGGATAGACCACGTACGTGTGATCA   2294

Qy   1441 TAACCACTTGGTACTACACACCCAGAAAACTCAAAATTGTCTTTCCTCCGATGAGATATGG   1500
          ||||||||||||||||||||||||||||||||||| ||||| ||||| |||||||||| |
Db   2295 TAACCACTTGGTACTACACACCCAGAAAACTCAAAAATTGTC-TTTCCCTGATGAGATAT-GG 2352

Qy   1501 GTGTCCTTTTGTACGTCTAGGGCCTAGGTGTACCCAAGTGAAGTTATATAAGCAAAT   1560
          ||||||||||||||||| || | |||| | |||||| ||| |||||||| ||||
Db   2353 GTGTCCTTTTGTACGTCAGGCC---TAGGTAACCAGTGGAGTGATTATTAGC-AAAT   2408

Qy   1561 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTT   1607
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   2409 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTT   2455

RESULT 8
AK125130
LOCUS   AK125130   3493 bp   mRNA   linear   PRI 09-SEP-2003
DEFINITION Homo sapiens cDNA FLJ43140 fis, clone CTONG3007528.
ACCESSION   AK125130
VERSION   AK125130.1 GI:34531119
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
            Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
            Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 3493)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-JUL-2003) Takao Isogai, FLJ Project(HRI Team); 2-6-7

```
LOCUS        AX332151              580 bp    DNA        linear   PAT 09-JAN-2002
DEFINITION   Sequence 2660 from Patent WO0194629.
ACCESSION    AX332151
VERSION      AX332151.1  GI:18122785
KEYWORDS     .
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1
  AUTHORS    Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
             Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE      Cancer gene determination and therapeutic screening using signature
             gene sets
  JOURNAL    Patent: WO 0194629-A 2660 13-DEC-2001;
             Avalon Pharmaceutical (US)
FEATURES             Location/Qualifiers
     source          1..580
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match             29.3%;  Score 475;  DB 6;  Length 580;
Best Local Similarity   95.7%;  Pred. No. 2.6e-81;
Matches  563;  Conservative    0;  Mismatches   16;  Indels   9;  Gaps   7;

Qy    2440 ATATGATTTACTGCTTGAAATTACATTGAGTTGAAGTTTAGAAACTAACATAGCATTA 2499
Db     601 ATATGATTTACTGCTTGAAATTACATTGAGTTGAAGTTTAGAAACTAACATAGCATTA 660

Qy    2500 ATATGAATAAT-CATGGAAAATTATATCCTTTGAAAACTAACAAGTTCAAACAGTTAGCAA 2558
           ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    2500 ATATGAATAAT-CATGGAAAATTATATCCTTTGAAAACTAACAAGTTCAAACAGTTAGCAA 2558

Qy     661 CTCCTTTAGAAACAGTCAAAGCCACTCAAACAGTTCAAACAGTTCAAACAGTTCAAACAGGAAGGTAGCAA 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2559 CTCCTTTAGAAACAGTCAAAGCCACTCAAACAGTTCAAACAGTTCAAACAGGAAGGTAGCAA 2618

Qy     721 GTTAGGGCGGATGATTATATTTCTTGGCTTGTGTATACCATTGGCCAGGGCCTTTATA 780
           ||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    2619 GTTAGGGCGGATGATTATATTTCTT-GCTTGTGTATA-CCAGTTGTCAAGGACATTATA 2676

Qy     781 AGGACTCCCAAAGCATTTGAAGAATGGCAATATCAAATAAGTGTATGTCCTCTCAAAT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2677 AGGACTCCCAAAGCATTTGAAGGATGGCAATATCAAATAAGTGTATGTCCTCTCAAAT 2736

Qy     841 GAGGCATTTTAATTGTACAATCTATTTGGACGCTCAGGTTATGATATGTTTATGAAAA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2737 GAGGCATTTTAATTGTAAAATCTATTTGGACGCTCAGGTTATGATATGTTTATGAAAA 2796

Qy     901 ATAAGCTTCATTATTCTTATAGCTACATCCTATTATTCCCTTTAGAAACAAGAATAAC 960
           ||||||||| ||||| |||||||||||||||||||||||||||||||||||||||||
Db    2797 ATAAGCTTCATTATTT-TTATAGCTACATCCTATTAT-CCCTTTAGAAACAAGAATAAC 2855

Qy     961 AATAAGTTTAATAGTTGCCATACTTAGCATTTATCAGTCTCAATGAAACCAATATTGAA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2856 AATAAGTTTAATAGTTGCCATACTTAGCATTTATCAGTCTCAATGAAACCAATATTGAA 2915

Qy    1021 TCTCTGATAAATATTTCTGATGTTACTAGCTATAGCTTTTCGGCACGGCACAACCCTG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2916 TCTCTGATAAATATTTCTGATGTTACTAGCTATAGCTTTTCGGCATGGCACAACCCTG 2975

Qy    1081 ACATTACTAAGTGGAAATGTTAGGATTTTCGGCATGCGTAGAATCTCTAAAATTT 1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2976 ACATTACTAAGTGGAAATGTTAGGATTTTCGGCATTGCATGTTAGAATCTCTAAAATTT 3035

Qy    1141 AAACATTCCTGTAAATGACTAAGGTTTGCTTTTATCAATATGAAATTCTGAAGGCCAATA 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3036 AAACATTCCTGTAAATGACTAAGGTTTGCTTTTATCAATATGAAATTCTGAAGGCCAATA 3095

Qy    1201 TCATACCATTAACTAACATGAAAGCTTTAATTCCTAAAAAATAGTTTAGACAATATTCAAGCA 1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3096 TCATACCATTAACTAACATGAAAGCTTTAATTCCTAAAAAATAGTTTAGACAATATTCAAGCA 3155

Qy    1261 ATGCTCCTAATAATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCTGGTTTA 1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    3156 ATGCTCCTAATAATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCT-GTTTA 3214

Qy    1321 AAAATGAAATCTTTATATTGACTGGGTGTCCCACATATTCCAGTAATTTCTGTTATGAGA 1380
           ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    3215 AAAATGAAATCTTTATATTGACT-GGTGTTCCACATATTCCAGTAATTTCTGTTATGAGA 3273

Qy    1381 GGACTTGAAATAGCAAATTGCCCACACAGTTAACTGGATCACGTACGTGGTGATCA 1440
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    3274 GGACTTGAAATAGCAAATTG-CCACACAGTTAACTGGATAGCACGTACGTGGTGATCA 3332

Qy    1441 TAACCACTTGGTACTACACCCGAGAAACTCAAAATTGTCTTTCCTGTAATATAAGCAAAAT 1500
           ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    3333 TAACCACTTGGTACTACACCCGAGAAACTCAAAATTGTC-TTTCCTGTAATATAAGC-AAAT 3390

Qy    1501 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAATTTATTTGGG 1560
           ||||||||||||||||||||||||||||||||||||||||||||||| 
Db    3391 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAATTTATTTTAATT 3446

Qy    1561 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAATTTATTT 1607
           |||||||||||||||||||||||||||||||||||||||||||
Db    3447 GTGTTTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAATTTATTT 3493

RESULT 9
AX332151/c
```

```
RESULT 10
AX334874/c
LOCUS        AX334874              580 bp    DNA        linear   PAT 09-JAN-2002
DEFINITION   Sequence 5383 from Patent WO0194629.
ACCESSION    AX334874
```

VERSION   AX334874.1 GI:18125593
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
TITLE   Cancer gene determination and therapeutic screening using signature
            gene sets
JOURNAL   Patent: WO 0194629-A 5383 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES   source
            Location/Qualifiers
            1..580
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN

Query Match   29.3%; Score 475; DB 6; Length 580;
Best Local Similarity   95.7%; Pred. No. 2.6e-81;
Matches   563; Conservative   0; Mismatches   16; Indels   9; Gaps   7;

Qy   1025 TGATAAATATTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACAT 1084
           ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   580 TGATANATATTTCTGATGTTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACAT 521

Qy   1085 TACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGAATCTCTAAAATTTAAAC 1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   520 TACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGAATCTCTAAAATTTAAAC 461

Qy   1145 ATTCCTGTTAAATGACTAAGGTTTGCTTTATCAATATGAATTCTGAAGGCCAATATCAT 1204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   460 ATTCCTGTTAAATGACTAAGGTTTGCTTTATCAATATGAATTCTGAAGGCCAATATCAT 401

Qy   1205 ACCATTAACTATGAAAGCTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCAATGC 1264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   400 ACCATTAACTATGAAAGCTTTAATTCCTAAAAATAGTTTTAGAGATATTCAAGCAATGC 341

Qy   1265 TCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCTGGTTTAAAAA 1324
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   340 TCTCCTAATATCCATACGCAAGTGTGTTATGACACAAATTCACTAGTCT-GTTTAAAAA 282

Qy   1325 TGAAATCTTTATATTGACTGGGTGTCCCACATATTCCAGTAATTTCGTTATGAGAGAC 1384
           |||| ||||||||||||| || ||| ||||||||||| ||||||||||||||||||||
Db   281 TGAATTCTTTATATTGACT-GGTGTTCCACATATTCAGTAATTTCGTTATGAGAGAC 223

Qy   1385 TTGAAATAGCAAATTGCCCACACAGTTAACTGGATAGATCACGTACGGTGATCATAAC 1444
           ||||||||||||||| ||||||||||||||||||||| ||||||||||| |||||||||
Db   222 TTGAAATAGCAAATTG-CCACACAGTTAACTGGATAGACCACGTACGTGGTGATCATAAC 164

Qy   1445 CACTTGGTACTACACCCAGAAACTCAAAATTGTCTTTCTCCTGATGAGATATGGGGTGT 1504
           |||||||||||||||||||||||||||||||||| |||||||||||||||| |||||
Db   163 CACTTGGTACTACACCCAGAAACTCAAAATTGTC--TTTCTCCTGATGAGATAT-GGGTGT 106

Qy   1505 CCTTTTGTACGTCTAGGGCCTAGGGTACCCAAGTGAATAATAATAAGCAAAATGTGT 1564
           ||||||||||||||||||||| |||| |||| |||| ||| ||||| ||||||||
Db   105 CCTTTTGTACGTCTAGGGCC---TAGGTAACCAGTGGAGTGATTATATTAGC-AAATGTGT 50

Qy   1565 TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAA 1612
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   49 TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTTATTTAAAAA   2

RESULT 11
AX335287/c
LOCUS   AX335287   580 bp   DNA   linear   PAT 09-JAN-2002
DEFINITION Sequence 5796 from Patent WO0194629.
ACCESSION   AX335287
VERSION   AX335287.1 GI:18126006
KEYWORDS   .
SOURCE   Homo sapiens (human)

RESULT 12
AX335859/c
LOCUS   AX335859   580 bp   DNA   linear   PAT 09-JAN-2002
DEFINITION Sequence 6368 from Patent WO0194629.
ACCESSION   AX335859
VERSION   AX335859.1 GI:18126578
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1
AUTHORS     Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
TITLE       Cancer gene determination and therapeutic screening using signature
            gene sets
JOURNAL     Patent: WO 0194629-A 6368 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1..580
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match         29.3%;  Score 475;  DB 6;  Length 580;
Best Local Similarity   95.7%;  Pred. No. 2.6e-81;
Matches  563;  Conservative   0;  Mismatches  16;  Indels   9;  Gaps    7;

Qy       1025 TGATAAATATTTTCTGATGTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACAT 1084
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 TGATANATATTTTCTGATGTACTAGCTATGGGAAATTAGAACTGGCACAACCCTGACAT 521

Qy       1085 TACTAGTGGAAATGTTAGGATTTTCGGCATCGCATCGTTAGAATCTCTAAAATTTAAAC 1144
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 TACTAAGTGGAAATGTTAGGATTTTCGGCATCGCATCGTTAGAATCTCTAAAATTTAAAC 461

Qy       1145 ATTCCTGTAAATGACTAAGGTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCAT 1204
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 ATTCCTGTAAATGACTAAGGTTGCTTTTATCAATATGAATTCTGAAGGCCAATATCAT 401

Qy       1205 ACCATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTAGAGATATTCAAGCAATGC 1264
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 ACCATTAACTATGAAAGCTTTTAATTCCTAAAAATAGTTTAGAGATATTCAAGCAATGC 341

Qy       1265 TCTCCTAATATCCATACGGCAAGTGTGTTTATGACACAAATTCACTAGTCTGGTTTAAAA 1324
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 TCTCCTAATATCCATACGGCAAGTGTGTTTATGACACAAATTCACTAGTCT-GTTTAAAA 282

Qy       1325 TGAAATCTTTATTGACTGGGTCCCACATATTCCAGTAAITTCTGTTATGAGAGGAC 1384
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 TGAATTCTTTATTGACT-GGTGTTCCACATATTCAGTAATTTCTGTTATGAGAGGAC 223

Qy       1385 TTGAAATAGGCAAATGCCCACACAGTTAACTGGATAGATCACGTACGTGGTGATCATAAC 1444
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 TTGAAATAGGCAAATTG-CCACACAGTTAACTGGATAGAGACCACGTACGTGGTGATCATAAC 164

Qy       1445 CACTTGGTACTACACCCAGAAAACTCAAAATTGTCTTTCTCCTGATGAGATATGGGGTGT 1504
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 CACTTGGTACTACACCCAGAAAACTCAAAATTGTC--TTTCCTGATGAGATAT-GGGTGT 106

Qy       1505 CCTTTTGTACGTCTAGGGCCTAGGGGCCTAGGGCC---TAGGTAACCAGTGGAGTGAATATATAAGCAAAATGTGT 1564
              ||||||||||||||||||||||||||||          ||||||||||||||||||||||||||||||||||||||
Db        105 CCTTTTGTACGTCTAGGGCC---------------TAGGTAACCAGTGGAGTGATTATATTAGC-AAATGTGT 50

Qy       1565 TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTATTTAAAA 1612
              ||||||||||||||||||||||||||||||||||||||||||||||
Db         49 TTGTATCCAGAGTCTTCCTGTCATTGTAATAAAAAATTATTTAAAA 2


RESULT 13
AC111087/c
LOCUS       AC111087            231508 bp    DNA     linear   ROD 30-SEP-2003
DEFINITION  Mus musculus chromosome 9, clone RP23-152016, complete sequence.
ACCESSION   AC111087
VERSION     AC111087.9  GI:33300743
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 231508)
AUTHORS     Birren,B., Nusbaum,C. and Lander,E.
TITLE       Mus musculus chromosome 9, clone RP23-152016
```

```
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 231508)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,R., Lamazares,R.,
            Landers,T., Lehoczky,J., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,N., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 231508)
AUTHORS     Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (03-JUL-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 231508)
AUTHORS     Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., Pierre,N.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
```

```
JOURNAL     Submitted (29-JUL-2003) Whitehead Institute/MIT Center for Genome
REFERENCE   Research, 320 Charles Street, Cambridge, MA 02141, USA
            5 (bases 1 to 231508)
AUTHORS     Birren,B., Nusbaum,C., Lander,B., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,Y., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,X., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vo,X.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 29, 2003 this sequence version replaced gi:32441344.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence submissions@genome.wi.mit.edu
            ----------- Project Information
            Center project name: L22440
            Center clone name: 152_O_16

FEATURES            Location/Qualifiers
     source         1..231508
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /chromosome="9"
                    /map="9"
                    /clone="RP23-152O16"
                    /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature   1..2033
                    /note="wgs end extension
                    clone_end:SP6"
     unsure         complement(1..96)
                    /note="single clone coverage"
     unsure         complement(1573..1577)
                    /note="<30 qual SNGL region"
     unsure         complement(1815..1849)
                    /note="single clone coverage"
     repeat_region  complement(2101..2522)
                    /rpt_family="MLT1G1"
     repeat_region  complement(2551..2714)
                    /rpt_family="MLT1G1-int"
     unsure         2592..2596
                    /note="<30 qual SNGL region"
     unsure         2738..2834
                    /note="single clone coverage"
     unsure         3092..3096
                    /note="<30 qual SNGL region"
     repeat_region  3100..3151
                    /rpt_family="(TTTG)n"
     unsure         3413..3418
                    /note="<30 qual SNGL region"
     unsure         3432..3436
                    /note="<30 qual SNGL region"
     repeat_region  3909..3933
                    /rpt_family="(TTTG)n"
     unsure         complement(4174..4190)
                    /note="single clone coverage"
     unsure         4291..4413
                    /note="single clone coverage"
     repeat_region  complement(4780..4869)
                    /rpt_family="MIR"
     repeat_region  5165..5339
                    /rpt_family="RSINE1"
     repeat_region  5355..5389
                    /rpt_family="(TAA)n"
     repeat_region  5394..5547
                    /rpt_family="B3A"
     repeat_region  5548..5608
                    /rpt_family="(CA)n"
     repeat_region  5609..5635
                    /rpt_family="B3A"
     repeat_region  complement(5636..5822)
                    /rpt_family="B2_Mm1"
     repeat_region  5823..5833
                    /rpt_family="B3A"
     repeat_region  5841..5947
                    /rpt_family="Lx8"
     repeat_region  6330..6355
                    /rpt_family="AT_rich"
     repeat_region  7348..7502
                    /rpt_family="ORR1D"
     repeat_region  8809..8905
                    /rpt_family="MLT1D"
     repeat_region  8906..9147
                    /rpt_family="MTE"
     repeat_region  complement(9151..9464)
                    /rpt_family="IAPLTR1_MM"
     repeat_region  9466..9479
                    /rpt_family="MTE"
     repeat_region  9480..9689
                    /rpt_family="RLTR10C"
     repeat_region  9690..9715
                    /rpt_family="(GA)n"
     repeat_region  9716..9916
                    /rpt_family="RLTR10C"


Query Match             7.3%;  Score 118.2;  DB 10;  Length 231508;
Best Local Similarity  55.0%;  Pred. No. 3.1e-13;
Matches  556;  Conservative  0;  Mismatches 368;  Indels  87;  Gaps  13;

Qy    225 TAATTGAAAAGGGATATAAAGCATTAAAATAATATAGCTATATATAGCTATATCATGTATGTA  284
          ||||| |||| ||  |||  |||||||| |||| | ||||||| | ||||| ||| |||||||
Db  66685 TAACTCATAGAAAGATGGAATTAAAGATATTAAAGATATATATACATGTGTATGTACACATATGTA 66626

Qy    285 TCTAACAGAGAAGTTCAAGTCACTCACTTGGAGCATGGGACCAGCC  344
          | | | | | ||||||| ||  || | |||    || | | | |||
Db  66625 TGTGTGTGTGTAGTTCATAATTTCATAAATTTTCTAAGCAACAGC 66566

Qy    345 TCAGGTCTTATGCTGGGATGCAGTAGACA-GGAGATGGGGAAAGAGAGAACTGT  403
          |  | | | | | |   |||| |  |  |  | ||| |    ||| |||||||
Db  66565 CTGAGATCCTTGTTGAAATGCGGCAGTGCAGTGCACCCTTAAGGAACTTATATTAATTAATGATGATCTGAATA  66506

Qy    404 GTAATTGAAATGACGTGGGCTGCACCCTTAAGGAACTTATAATTAATTAATTATTGAGCTGGATG  463
          ||||| || |  | |||| || ||| | |||  || || | |||||| || ||| |||||||||
Db  66505 ATCTTTAAGCTGACTTGCCCTGTAAGCTTTGGGACTTGCAATTAATTATTGAGCTGGATG 66446

Qy    464 AACAT------ACCAGGATAAAGATGTCAAATGGTCAAATGAGTGTGACTCCCTAAAGTAGATTA  515
          |     |      | || |  | || | ||| |  |  |  |||| ||  | | | ||||||| |
Db  66445 AATGTGAATATGGATCTAATGAATGATCAAAAAACTGTAATGCCTTCAAAGTAGATGA 66386

Qy    516 AAGTGTGCATTCTTTGTTTCCTCAAAAATAATGATTTTACTGCTTGAAATTACATTTGAGTTG  575
          || |||| ||||||| ||||||| ||||||| | || | || ||   || ||||| || |||
Db  66385 AAATATATATACTTTCATGTTTTCCCAAAAAATAGCATGTCTCTGCATGAAGTCAATTTTGATTTG 66326

Qy    576 AAGTTTAGAAACTAACATAGCCATTAATGAATGCATGGAGCATGCGAAAATTTATATCCTTTGA  635
          || |||  |  |||| |||  |  |||||||| ||| ||||| | ||||| |||||||| | |
```

```
Db  66325  TA-------------CTTAAACTGTAATATGTGTGTTGTTAATAGTAATACCCATGGA 66285
                          ||| ||||||||  ||  | || ||||| ||||||||||||||
Qy    636  AAAACTGATTGATAAATATATTCCCCTCCTTTAGAAACAGTCAAAAGCCACTTCAAACAA 695
Db  66284  AACCATTACTCTTCTGAACATTTACCT-GTTTGATATGGCTATGAGTCATTTGAAGCAA 66226
           ||   | |  |||||  |||||  |||| |||||| |||  |||||| |||| |||||||
Qy    696  GTTTCAAAATAAAGGAAGGTAGCAAGTTAGGCGATGGATTATATTTTCTTGGCTTGTTGT 755
Db  66225  GTTACAAA---------CAAGTCAGGCAGGGGATAGATTCTGTCTCTT------------ 66189
           | || ||          |   |  |  | |||  |   |  | ||| |
Qy    756  ATACCCATTGGCCCAGGGCCTTTATAGGACTCCCAAAAGCATTTGAAGAATGGCAATAT 815
Db  66188  -----------------TCAAGGCCCATCACAAGCATTTTGGGTGAT-GAAGTAT 66153
                            ||||| ||||| |||||| ||| ||| | | |||||||
Qy    816  CAAATAAGTGTATGTCCCTCAAATGAGGCATTTTAATTGTTACAATCTATTTGGACGC 875
Db  66152  TAAATCAGTGCATGTTTCTCAACAG-GAAAGTTTAATAGTTAAATCCCAAGTACACATAT 66094
           |||| |||||  ||||| ||||||| |||||||||||||||||||| ||||||| || ||
Qy    876  TCAGGTTATGATATGTTTATGAAAAAATAAGCTTCATTATTTCTTATAGCTACATCCTATT 935
Db  66093  TAAAAGTCTGGTATGTTTATTAAAGTAAGTTTTATTATTT-TCATTGATATGTCCTATT 66035
           |||| ||| || |||||| || ||| |||||| |||||||  |||||||||| |||||||
Qy    936  ATTCCCTTTTAGAAACAGAGAATAACAATAAAGTTTAATAGTTGCCATACTTAGCAT-TTA 994
Db  66034  ATTCCGTTT----ACAGAGAATAACAGTATATTTAAATCATTGCTGTATTTAGCATATTT 65979
           ||||| ||||    ||||||||||| |||| |||| || ||| |  ||||||| ||| ||
Qy    995  TCAGGTCTAATGAAACCAATATGAATCTCTGATAAATATTTTCTGATGTTACTAGCTAT 1054
Db  65978  TCAGATCTAAAGAAACCATTAATATAAATCCTTGACAAATATTTCATGGAGTCTCCAGTGCT 65919
           |||| |||||| |||||| | |||  |||  ||||||||||| | ||| || |||||| ||
Qy   1055  GGGAAATTAGAACTGGCACAACCCTGACATTACTAAGTGGAAATGTTAGGATTTTTCGGC 1114
Db  65918  GGGAAACCAAACATGGTGGAAGTTGACTTAACTAACCAG-AGTTTTATAATTCTGGAGC 65860
           ||||| | |||| |||| || ||| ||| |||||| ||| ||||| | |||| ||||||
Qy   1115  ATGCATGTTAGAAT---CCTAAAATTTAAACATTCCTGTTAAATGACTAAGGTTTGCT 1171
Db  65859  ACCATATGTTTGAGTTCATATTAATTATATGAACAAAATTCATATTAAATGACTTCATTATGAC 65800
           | || | |||| ||||| |||| ||  || ||| ||||| ||||| ||||| || ||||| ||
Qy   1172  TTTATCAATAATGAATTCTGAAGGCCAATATCATTAACTATGAAAGC 1222
           ||| | |||||||| |  |||||||||    ||   |||
Qy   1222
Db  65799  TCTTTCTTATCATTTTCTAAAAGCCAACTACTATAAACAACAACATAAC 65749
```

RESULT 14
AC116714

LOCUS   AC116714   256138 bp   DNA   linear   HTG 23-APR-2003
DEFINITION Mus musculus clone RP23-122E19, WORKING DRAFT SEQUENCE, 35
            unordered pieces.
ACCESSION   AC116714
VERSION   AC116714.2 GI:30018030
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 256138)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 256138)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,P.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,A., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,A., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 256138)
  AUTHORS   Birren,B., Nusbaum,C., Lander,B., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,T., Maclean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Wu,X.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Apr 17, 2003 this sequence version replaced gi:19881957.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ---------------- Project Information ----------------
            Center project name: L22412
            Center clone name: 122_E19
            ----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 35 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *   1: contig of 57828 bp in length
          *   57828: gap of 100 bp
          *   57829: contig of 624 bp in length
          *   58552: gap of 100 bp
          *   58553: contig of 27 bp in length
          *   58653: gap of 100 bp
          *   58679: contig of 100 bp
          *   58780: gap of 100 bp
          *   58779: contig of 620 bp in length
          *   59339: gap of 100 bp
          *   59400: contig of 998 bp in length
          *   59499: gap of 100 bp
          *   59500: contig of 998 bp in length
          *   60497: gap of 100 bp
          *   60498: contig of 100 bp
          *   60597: contig of 677 bp in length
          *   60598: gap of 100 bp
          *   61274: gap of 100 bp
          *   61275: contig of 100 bp
          *   61374: gap of 100 bp
          *   61375: contig of 977 bp in length
          *   62351: gap of 100 bp
          *   62352: contig of 100 bp
          *   62451: gap of 100 bp
          *   62452: contig of 1038 bp in length
          *   63489: gap of 100 bp
          *   63490: contig of 100 bp
          *   63589: gap of 100 bp
```

```
*   63590   64252: contig of 663 bp in length
*   64253   64352: gap of 100 bp
*   64353   65345: contig of 993 bp in length
*   65346   65445: gap of 100 bp
*   65446   66388: contig of 943 bp in length
*   66389   66488: gap of 100 bp
*   66489   67203: contig of 715 bp in length
*   67204   67303: gap of 100 bp
*   67304   68016: contig of 713 bp in length
*   68017   68116: gap of 100 bp
*   68117   68822: contig of 706 bp in length
*   68823   68922: gap of 100 bp
*   68923   70247: contig of 1325 bp in length
*   70248   70347: gap of 100 bp
*   70348   71695: contig of 1348 bp in length
*   71696   71795: gap of 100 bp
*   71796   72936: contig of 1141 bp in length
*   72937   73036: gap of 100 bp
*   73037   74103: contig of 1067 bp in length
*   74104   74203: gap of 100 bp
*   74204   75393: contig of 1190 bp in length
*   75394   75493: gap of 100 bp
*   75494   77147: contig of 1654 bp in length
*   77148   77247: gap of 100 bp
*   77248   78921: contig of 1674 bp in length
*   78922   79021: gap of 100 bp
*   79022   80676: contig of 1655 bp in length
*   80677   80776: gap of 100 bp
*   80777   82008: contig of 1232 bp in length
*   82009   82108: gap of 100 bp
*   82109   83438: contig of 1330 bp in length
*   83439   83538: gap of 100 bp
*   83539   84840: contig of 1302 bp in length
*   84841   84940: gap of 100 bp
*   84941   86247: contig of 1307 bp in length
*   86248   86347: gap of 100 bp
*   86348   88250: contig of 1903 bp in length
*   88251   88350: gap of 100 bp
*   88351   90065: contig of 1715 bp in length
*   90066   90165: gap of 100 bp
*   90166   99884: contig of 9719 bp in length
*   99885   99984: gap of 100 bp
*   99985  111968: contig of 11984 bp in length
*  111969  112068: gap of 100 bp
*  112069  126928: contig of 14860 bp in length
*  126929  127028: gap of 100 bp
*  127029  145797: contig of 18769 bp in length
*  145798  145897: gap of 100 bp
*  145898  172246: contig of 26449 bp in length
*  172247  172346: gap of 100 bp
*  172347  218202: contig of 45756 bp in length
*  218203  218302: gap of 100 bp
*  218303  256138: contig of 37836 bp in length.
```

```
FEATURES             Location/Qualifiers
source               1..256138
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-122E19"
                     /clone_lib="RPCI-23 Female Mouse BAC"
misc_feature         1..57828
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
misc_feature         57929..58552
                     /note="assembly_fragment"
misc_feature         58653..58679
                     /note="assembly_fragment"
misc_feature         58780..59399
                     /note="assembly_fragment"
misc_feature         59500..60497
                     /note="assembly_fragment"
misc_feature         60598..61274
```

```
misc_feature     /note="assembly_fragment"
                 61375..62351
misc_feature     /note="assembly_fragment"
                 62452..63489
misc_feature     /note="assembly_fragment"
                 63590..64252
misc_feature     /note="assembly_fragment"
                 64353..65345
misc_feature     /note="assembly_fragment"
                 65446..66388
misc_feature     /note="assembly_fragment"
                 66489..67203
misc_feature     /note="assembly_fragment"
                 67304..68016
misc_feature     /note="assembly_fragment"
                 68117..68822
misc_feature     /note="assembly_fragment"
                 68923..70247
misc_feature     /note="assembly_fragment"
                 70348..71695
misc_feature     /note="assembly_fragment"
                 71796..72936
misc_feature     /note="assembly_fragment"
                 73037..74103
misc_feature     /note="assembly_fragment"
                 74204..75393
```

```
Query Match        7.3%;   Score 118.2;   DB 2;   Length 256138;
Best Local Similarity  55.0%;   Pred. No. 3-13;
Matches  556;  Conservative  0;  Mismatches  368;  Indels  87;  Gaps  13;

Qy      225 TAATTGAAAAAGGATATAAAGCATTAAATAATATGAGATAATATAGCTATATCTATGTATGTA 284
                 |||||||  |||  ||||||||||||| | | ||||  |||| | || |||||| ||||||
Db    16813 TAACTCATAGAAAGATGGAATTAAAGATATTAAAGATATATATATATACATGTATGTACACATATGTA 16872

Qy      285 TCTAACAGAGAAGTCAAGTCACTTCAATTAAATGAAACATTTTGAGCATGGACCAGCC 344
                 |||||||||  ||||||||||||||||| ||||||||||| |||||||||||||
Db    16873 TGTGTGTGTGTAGTTCATATAATTTCATAAAATTTTAATTCTAATTTTCTAAGCAACAGC 16932

Qy      345 TCAGGTCTTATGCTGGGATGCAGTCAGTAGACA-GGAGATGGGGAAATTAGAAAAAGAAGAACTGT 403
                 |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||
Db    16933 CTGAGATCCTTGTTGTTGAAATGCGGCAGTCAGCCCTTAAGGAACTTATAATTAATGATGATCTGAATA 16992

Qy      404 GTAATTGAAATGACGTGGGCTGCACCCTTAAGGAACTTATAATTAATGATGATCTGAATA 463
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16993 ATCTTTAAGCTGACTGCCCGTAAGCTTGGGACTTGCAATTAAATTATTGAGCTGGATG 17052

Qy      464 AACAT------ACCAGGATAAAGATGTCAAATGAGTGGACTCCCTTAAAGTAGATTA 515
                 |||||      ||||||||||| ||||||||||||||||||||||||| |||||||||
Db    17053 AATGTGAATATGGATCTCAATGATGAATATCAAAAAACTGTAATGCCTTCAAAGTAGATGA 17112

Qy      516 AAGTGTGCATTCTTTGTTTCCTAAAAATGATTTTACTGCTTGAAATTACATTTGAGTTG 575
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    17113 AAAATATCTTTCATGTTTTCCCAAAATAGCATGTCTCTGCATGGAAGTCAATTTGATTTG 17172

Qy      576 AAGTTTAGAAACTAACATGACATTAAAATGAATGATGAGAAAATTATTATCCTTTGA 635
                 ||||||||||||||||||                             || |||||||
Db    17173 TA------------------CTTAAACTGTAATATGTTGTTAATAGTAATAACCCATGGA 17213

Qy      636 AAACTGATTGATAAATATATTCCCCCTCCTTTGAAACAGTCAAAAGCCACTTCAAACAA 695
                 ||||||||||||||||||                          | ||||||||||||
Db    17214 AACCATTATCTTCTGAACATTTTACCT-GTTTGGATATGGCTATGAGTCATTTGAAGCAA 17272

Qy      696 GTTTCAAAATAAAGAAGGTAGCAAGTAGGCGATGGATTAGATATATTTTCTTGGCTTGTTGT 755
                 |||||||||                                ||||||| ||||||
Db    17273 GTTACAAA--------CAAGTCAGGCAGGGGATAGATTCTGTCTT------------ 17309

Qy      756 ATACCCATTGGCCAGGGCCCTTTATAAGGACTCCCAAAGCATTTGAAGAATGGCAATAT 815
                                                    ||||| | ||||||||||||||
Db    17310 ------------------------------TCAAAGCCCATCACACAGCATTTTGGGTGTGAT-GAAGTAT 17345

Qy      816 CAAATAAGTGTATGTCCCTCAAATGAGGCATTTTAATTGTTACAATCTATTTGGACGC 875
```

```
Db  17346 TAAATCAGTGCGATGTTTCTAACAAG-GAAAGTTTAATAGTAAATCCCAAGTACATAT 17404
Qy    876 TCAGGTTATGATATGTTTATGAAAAATAAGCTTCATTATTTCTTATAGCTACATCCTATT 935
Db  17405 TAAAAGTCTGGTATGTTTATTAAAGTAAGTTTTATTATTT-TCATTGATATGTCCTATT 17463
Qy    936 AATTCCCTTTAGAAACAAGAATAACACATAAGTTTAATAGTTGCCATACTTAGCAT-TTA 994
Db  17464 ATTCCGTTT---ACAGAGAATAACAGTATATTTAAATCATTGCTGTATTTAGCATATTT 17519
Qy    995 TCAGGTCTAATGAAACCAATATTGAATCTCTGATAAATATTTTCTGATGTTACTAGCTAT 1054
Db  17520 TCAGATCCAAAGAAGAAACCATTAAATAAATCCTTGACAAATATTTCATGGAGTCTGCAGTGCT 17579
Qy   1055 GGGAAATTAGAACTGGCACAACCCTGACATTACTAAGTGGAAATGTTAGGATTTTTCGGC 1114
Db  17580 GGGAAACCAAACATGGTGGAAGCTTGACTTAACTAACCAG-AGTTTTATAATTCTGGAGC 17638
Qy   1115 ATCGCATGTTAGAAT--CTCTAAAATTTAAACATTCCTGTTAAAATGACTAAGGTTTGCT 1171
Db  17639 ACCATATGTTTGAGTTCATATTATTAATATGAACAAATTCATATTAAATGACTTCATTATGAC 17698
Qy   1172 TTTATCAATATGAATTCTGAAGGCCAATATCATACCATTAACTATGAAAGC 1222
Db  17699 TCTTTCTTTATCATTTTCTAAAAGCCAACACTACATAAACAACACAATAAC 17749
```

```
RESULT 15
AC132694/c
LOCUS       AC132694    202314 bp    DNA    linear    HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-490C22, WORKING DRAFT SEQUENCE.
ACCESSION   AC132694
VERSION     AC132694.3  GI:25007304
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 202314)
AUTHORS     Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aovagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswaio,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,J., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,C., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,A., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Heiderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., McNeill,T.Z., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montenayor,J., Moore,S.,
            Morgan,M., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu.L-L.,
            Puazo,M., Quiroz,J., Rachlin,J., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Wilson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,X., Zhou,J., Zhao,S., Dunn,D., von
            Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 202314)
AUTHORS     Rat Genome Sequencing Consortium.
TITLE       Direct Submission
JOURNAL     Submitted (03-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 202314)
AUTHORS     Rat Genome Sequencing Consortium.
TITLE       Direct Submission
JOURNAL     Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15,2002 this sequence version replaced gi:23908401.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            --------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu
            Contact: hgsc-help@bcm.tmc.edu
            --------- Project Information
            Center project name: KCCJ
            Center clone name: CH230-490C22
            --------- Summary Statistics
            Assembly program: PhraP; version 0.990329
            Consensus quality: 169128 bases at least Q40
            Consensus quality: 170817 bases at least Q30
            Consensus quality: 171921 bases at least Q20
            Estimated insert size: 173670; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 1 contigs. Gaps between the contigs
            *   are represented as runs of N. The order of the pieces
            *   is believed to be correct as given, however the sizes
            *   of the gaps between them are based on estimates that have
            *   provided by the submitter.
            * This sequence will be replaced
            *   by the finished sequence as soon as it is available and
            *   the accession number will be preserved.
            *   1  202314: contig of 202314 bp in length.
FEATURES             Location/Qualifiers
     source          1..202314
```

```
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-490C22"
     misc_feature   1..2170
                    /note="wgs_contig"
     misc_feature   2221..3503
                    /note="wgs_contig"
     misc_feature   complement(199582..200007)
                    /note="clone_boundary
                    clone_end:T7
                    site:
                    end_sequence:B2125806"
     misc_feature   201299..202314
                    /note="wgs_end_extension
                    clone_end:T7"

ORIGIN

Query Match        5.9%;   Score 96.2;  DB 2;   Length 202314;
Best Local Similarity  55.7%;  Pred. No. 5.3e-09;
Matches  492;  Conservative  0;  Mismatches  313;  Indels  78;  Gaps  13;

Qy    262 ATATAGCTATATCTATGCTATGTATGTATCTAACAGAGAAGTTCAAGTCACTTCA-ATTAAAGAA 320
          ||||||||  |||  |   |  ||||| | |||||  ||| || ||||||||||||  ||||||||
Db  14219 ATACAGGTATATACACATATGTATGTGTGTGTGTTGTGTTCATATAATTCATACGATTTTA 14160

Qy    321 ACATTTTGAGACATGGGACCAGCCTCAGGTCTTATGCTGGGATGGCAGTAGACA-CGGAGAT 379
          |||||  ||||||||| ||  ||||| |  | ||||| | | ||| || | || | ||||||
Db  14159 ATTCTGCTTTTCTGAGCATCAGCCTCAGATCCTGTGTTTAGGATTCAGCAATCAGCGGGGAA 14100

Qy    380 GGGGAAATTAGAAAAGAAGAACTGTGTAATTGAAATGACGTGGGCTGCACCCTTAAGGAAC 439
          ||   || ||||  ||||  |  ||||||| ||| | ||||  | | |||  |||||||||
Db  14099 GGAGAGATTCAAAAGGAGAGAAAATACATCTTCGAACTGACTTTGCCCCTGTAACCTTTGGGACT 14040

Qy    440 TTATAATTAAATGATGATCTGAATAAACATACCAGGATAAAGATGTCAAATGAGTGTGACT 499
          || || ||   |||| |   ||||||  |   ||  || ||  |  |||| |  | ||| |
Db  14039 TAAAATTAAATATTGAGTTG--GAAGAATCATCTAATATGAAATCAAAACCTGTAATG 13982

Qy    500 CCCTTAAAGTGAGATTAAAGTGCAATTCTTTGTTTCCTAAAATATGATTTTACTGCTTGA 559
          |||| |||    ||  |||||  || |||| | ||| | | ||||  || | |  | |||
Db  13981 CCTTCAAAGTAGAATGAAAATATCCTTC-ATGTTTCCTAAAAATAGCAGGTCTTTTGCATGA 13923

Qy    560 AATTACATTTGAGTTGAAGTTTAGAAACTAACATACACTAATATGAATAATGACATGGAA 619
          ||||||||  |||| ||  ||||||   |  || | |  | ||| || || ||||||||
Db  13922 AGTCAATTTGATTTGTACTT---AAACTGTAATATGTTGTTAACAGTAATAACCATGGA 13866

Qy    620 AATTATTATCCTTGAAAACTGATTGATAAATATATTCCCCTCCTTAGAAACAGTCAA 679
          |||||||||     | || |    |           |||| ||  || ||||| ||
Db  13865 AACCATTATCTTCTGAACAC-----------------GTTACCTGTTTGGATATGCGTAT 13823

Qy    680 AAGCCCACTTCAAACAAGTTTCAAAATAAAGGAGGTAGCAAGTTAGGCGATGGATTATAT 739
          |||  | |||| || |||  | |  |||||||    | |    |||  || ||  ||||||
Db  13822 AAGTTATCTGAAGCAAGTTAGAA---------------CAAGTCAGGCAGTGGGATAGAT 13777

Qy    740 TTTCTTGGCTTGTGTATACCCCATTGGCCAGGCCTTATAAAGGACTCCCAAAAGCATTT 799
          |||                              ||||| | |||  |  ||||||||||
Db  13776 T------------------------------------GTGGTTTTCAAGGCCATCACAAGCCATTT 13748

Qy    800 TGAAGAATGGCAATATCAAATAAGTGTATGTCCTCCAAATGAGGCATTTTAATTGTTA 859
          |  |||   | ||  ||||| ||||  ||  |||     |  ||  ||||| ||||||
Db  13747 TG-TGTGGCGAAAATATAAAATAAGTAGTGATGTCTTTCAA----TGGAAAATTTTAATAGTT 13693

Qy    860 CAATCTATTTGGACGCGTCAGGTTATGCGATATGTTATGAAAAATAAGCTTCATTATTTCTT 919
          ||| |||| |||| | |    | |||  | | || |||||||  ||  | |||| |||||||
Db  13692 AAACCCAAGTACATATTCTAAGTCTGGTCTGTTCTGTTTATTAAAAAAGTGAGTTTTATTAATT-TC 13634

Qy    920 ATAGCTACATCCTATATTCCCTTTAGAAACAGGATAACAATAAGTTTAATAGTTGC 979
          |   |   | | ||| |     | || || | | |    | | || ||||| |||
Db  13633 ATTGATATGACCCATTATTCCC-CTTACACAAAACAATAACAACATATCTGAATCGGCTGC 13575

Qy    980 CATACTTAGCATTTATCAGGTCAATGAACCAATATTGAATCTCTGATAAAATATTTTCT 1039
          ||| | || || | || ||| || |  || ||  | |||| | ||||  ||| || |||
Db  13574 TGTATTAGCATAATTCGAGATCTAAAGAAGAAAATTAATAAAATCCTTGACAAATATTTCAT 13515

Qy   1040 GATGTTACTAGCTATGGGAAAATTAGAACTGGCACAACCCTGACATTACTAAGTGGAAATG 1099
          ||| | ||| |||||| |||||||||| ||||  | | || | |||| | || |||||||
Db  13514 GGTGTCAGTAGTGCTGCGGGAAATCAGATATGGTGGAACCTTTACTTACCTAACTATCCAG-AGTT 13456

Qy   1100 TTAGGATTTTTCGGCATGCATGTTAGAATCTCTAAAATTTAA 1142
          ||| || | || |||||| ||  |||| ||| |  | |||||
Db  13455 TTATGATTCTAGATCACCATAAGTCTGAGGTACTACAAATTCA 13413


Search completed: April 25, 2004, 18:58:14
Job time : 4272.17 secs
```

```
                GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:     April 25, 2004, 07:00:48 ; Search time 1262.83 Seconds
                                        (without alignments)
                            16458.288 Million cell updates/sec

Title:       US-10-051-835-14
Perfect score: 696
Sequence:    1 taaaatacctgataggcagt........ttttgaaaatatataaaat 696

Scoring table: IDENTITY_NUC
            Gapop 10.0 , Gapext 1.0

Searched:    27513289 seqs, 1493109276 residues

Total number of hits satisfying chosen parameters:    55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
            Listing first 45 summaries

Database :   EST:*
         1:  em_estba:*
         2:  em_esthum:*
         3:  em_estin:*
         4:  em_estmu:*
         5:  em_estov:*
         6:  em_estpl:*
         7:  em_estro:*
         8:  em_htc:*
         9:  gb_est1:*
        10:  gb_est2:*
        11:  gb_htc:*
        12:  gb_est3:*
        13:  gb_est4:*
        14:  gb_est5:*
        15:  em_estfun:*
        16:  em_estom:*
        17:  em_gss_hum:*
        18:  em_gss_inv:*
        19:  em_gss_pln:*
        20:  em_gss_vrt:*
        21:  em_gss_fun:*
        22:  em_gss_mam:*
        23:  em_gss_mus:*
        24:  em_gss_pro:*
        25:  em_gss_rod:*
        26:  em_gss_phg:*
        27:  em_gss_vrl:*
        28:  gb_gss1:*
        29:  gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result        %     Query
 No.   Score  Match  Length  DB  ID          Description
-----  -----  -----  ------  --  --          -----------
c  1   262.4  37.7    286    10  AW838776    AW838776 CM4-LT005
c  2    67.6   9.7    997    29  CNS005TE    AL060767 Drosophil
c  3    61.2   8.8    987    29  CNS00418    AL066537 Drosophil
c  4    55.4   8.0   1101    29  CNS0039G    AL063921 Drosophil
```

```
c   5   54.8  7.9   846  28  B18548       B18548 T408-Sp6 TA
c   6   50.8  7.3  1373  29  CNS01637     AL106237 Drosophil
c   7   49.4  7.1  1101  29  CNS000D1     AL065414 Drosophil
c   8   49.4  7.1  1101  29  CNS017RP     AL108415 Drosophil
c   9   48.6  7.0  1101  29  CNS0181N     AL108773 Drosophil
c  10   48.2  6.9   687  29  CE833588     CE833588 tigr-gss-
c  11   48.2  6.9   878  29  CE833588     AL108993 Drosophil
c  12   47.8  6.9  1478  12  BM414951     BM414951 OP20019 M
c  13   47.6  6.8   932   9  AL514901     AL514901 AL514901
c  14   47.6  6.8   994  13  BX414650     BX414650 BX414650
c  15   47.4  6.8  1201  13  BX461310     BX461310 BX461310
c  16   46.8  6.7  1101  29  CNS0039L     AL063926 Drosophil
c  17   46.6  6.7   986  14  CD387197     CD387197 AGENCOURT
c  18   46.6  6.7  1092  13  BX359398     BX359398 BX359398
c  19   46.6  6.7  1201  13  BX360615     BX360615 BX360615
c  20   46.6  6.7  1201  13  BX387747     BX387747 BX387747
c  21   46.4  6.7  1101  29  CNS0039R     AL063932 Drosophil
c  22   46.4  6.7  1201  13  BX375239     BX375239 BX375239
c  23   46.2  6.6   617  28  AZ077015     AZ077015 RPCI-23-4
c  24   46.2  6.6   854  12  BM414967     BM414967 OP20035 M
c  25    46   6.6  1101  29  CNS00F86     AL070260 Drosophil
c  26   45.8  6.6   976  13  BX411173     BX411173 BX411173
c  27   45.8  6.6  1109  14  CA453621     CA453621 AGENCOURT
c  28   45.6  6.6   970  13  BX349797     BX349797 BX349797
c  29   45.6  6.6  1101  29  CNS0182P     AL108811 Drosophil
c  30   45.4  6.5   865  12  BG343667     BG343667 HVSMEg000
c  31   45.4  6.5  1201  13  BX361080     BX361080 BX361080
c  32   45.4  6.5  1209  29  CG749482     CG749482 P043-3-H1
c  33   45.2  6.5  1201  13  BX385572     BX385572 BX385572
c  34   44.8  6.4   982  29  CNS016V1     AL107239 Drosophil
c  35   44.8  6.4  1035  29  CNS04JPG     AL293821 Tetraodon
c  36   44.8  6.4  1101  29  CNS0006J     BX439822     BX439822
c  37   44.8  6.4  1101  29  CNS0006J     AL062049 Drosophil
c  38   44.8  6.4  1101  29  CNS017KX     AL108171 Drosophil
c  39   44.8  6.4  1201  13  BX355642     BX355642 BX355642
c  40   44.6  6.4   313   9  AA370340     AA370340 EST82022
c  41   44.6  6.4   954  13  BQ670925     BQ670925 AGENCOURT
c  42   44.6  6.4   954  13  BX326810     BX326810 BX326810
c  43   44.6  6.4  1028  13  BX431780     BX431780 BX431780
c  44   44.6  6.4  1201   9  AL538931     AL538931 AL538931
c  45   44.6  6.4  1201  13  BX361533     BX361533 BX361533
```

ALIGNMENTS

```
RESULT 1
AW838776/c
LOCUS       AW838776               286 bp    mRNA    linear    EST 18-MAY-2000
DEFINITION  CM4-LT0057-140100-080-h05 LT0057 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW838776
VERSION     AW838776.1  GI:7932750
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
   MEDLINE   20202663
    PUBMED   10737800
  COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
```

collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Oscogawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
  source   Location/Qualifiers
              1..997
              /organism="Drosophila melanogaster"
              /mol_type="genomic DNA"
              /db_xref="taxon:7227"
              /clone="BACR12K22"
              /clone_lib="RPCI-98"
              /note="end : TET3"

ORIGIN

Query Match   9.7%; Score 67.6; DB 29; Length 997;
Best Local Similarity   17.2%; Pred. No. 0.0014;
Matches   113; Conservative 198; Mismatches 346; Indels 0; Gaps 0;

Qy   36 GGAACACAGGAGGAGAGGCTTTGATGGCGATACAGACTTGGAAGACATCAGTTGCTGAGCAG   95
           ::|| : |::||:||||: :| : :||| ::||||| ||||: :|||||| ||| |||
Db   992 RRAGARARARARRGARRRGRGARGARARRRARRRRARGARARAGARRAAARGAGARRAGARA 933

Qy   96 TAAATGAGATGATTCAGGAAGAGTATAAACTGGGAAGAGGACAGAGGACAGGCTCAAGG   155
           :|::|| :|| ||: :: |::   ::||| :::||||:: :::||:| ::||| |||||
Db   932 RAARRRRARRRRAGRRAGRGRAGGRGRSAGGGRGRRAAARRAARDARARRARARARARRRA 873

Qy   156 AACATATTTAAGGACTGGGTAGAAAAACAAGAGAGTATGAACAAGAGAGTGAGGAGATTATT   215
           : |::::: ::|::|| ::: ::||::|::|:|::|| :::| ::|||||
Db   872 RAKAARARGARRRGRRRRGRAGRRRRGRAGRRGRGRRRRARRRRGAARRARA 813

Qy   216 AGCAGTGACCTTTGAGAGTACATCTCTAGAGTGGTATGTGTAGAGCCAGATTTTCAGCAT   275
           ::: : :: ||::|::||||| |: ::| :::|: || || ||||
Db   812 RGRRARARARRRRGRRRRGARGARGRRRRGGRRGGRRGRRGRRRRARARRRARARRRR   753

Qy   276 CAGAACCATCAGCATTTTGGGGTGAAGGAAAAGGAGCCATGAACTCAAAGGTGGGGA   335
           :|: :| | :::|: :::||| :::::||::::|:||:|
Db   752 AGARAARRRARRAGAAARRARRRRRRRRAGAARARRARRRARRRARRARGAGARRRR   693

Qy   336 AAAGGCCTTTGGGAGGTGGCAGTGTGCGGTAGGAGTTCCGGTTAGAATACATGG   395
           :::: :::||| :|: |: ||:|:::| :||||:||
Db   692 GRRGRGAGARRRGRGRGRGRGARRRRGRGRMRTRARRRRRRARGAGARRRARGARR   633

Qy   396 ATGAAAGAAAGGTGAGCACAGCCATTCCTTTACACAGAAATAACAGATTCCCAGCCT   455
           |:|| : ||||| :| |: :::|:: :::||| :|::: |:::||| :::||:|
Db   632 RRGRGRGRRRGARGRRRGRGRRRRGRGRRRGARRGRRRRRRRARGAGARRGRR   573

Qy   456 CTAACCAAAGAAACAACAAGTTTGGGAACATTCCTCCTTCTGAAATATGAAAGAGAGGG   515
           ::|| :: |:: :::||| :|||:|:|: :::||| ||:: ||::|||| ::|||
Db   572 RRGRARGARRGAARRRRARARRRRRRAAARRGRAGRARARGRAGAAAAAAAAGAACAGCCCAAGT   513

Qy   516 GATAAATACTGGAGTAGGATTGTGAAAAAAGTCAAGAGAAAAAAARARRAGAAAGTC   575
           :: :::::|| ||:|:::::|:|::::| :::||:::|::::|::
Db   512 RRGGARRRGGRRCGRRRAAGARRRRARARARRRARARARRRAGRAGARRRRTNNNNNNN   453

Qy   576 GTAACAGATACTTCTCCATGGGATGGATGGTAAAAAAGGAAGTTATTTCAATAAAATGACCCCTT   635
           ::||: ||| |:|::::||| || || |:||::| |||::::::|||:: |: :||:|||:
Db   452 NNNNNNNNAGCNTNNNTNTNTNTNTNNNNNNNNNCATANNNTNTTTTNTTT   393

Qy   636 GGAAGGAGTTCAAAACAGGTTGCCATATGCTTTATGTGGAGTTTGAAAAATATATA   692
           :||| :|| || ||| || |::||| |: | ::|| || |:| |||::::|
Db   392 NTNTNANNNNTTTTTTTNTNNNTTTNNNTTTNNNTTNNNTTTANTNTNTNNNNTTTNANA   336

Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM4-LT0057-140
100-080-h05&t3=2000-01-14&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 4
High quality sequence stop: 286.

FEATURES
  source   Location/Qualifiers
              1..286
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /dev_stage="Adult"
              /clone_lib="LT0057"
              /note="Organ: leiomios; Vector: puc18; Site_1: SmaI;
              Site_2: SmaI; A mini-library was made by cloning products
              derived from ORSTES PCR (U.S. Letters Patent application
              No. 196,716 - Ludwig Institute for Cancer Research)
              profiles into the pUC 18 vector. Reverse transcription of
              tissue mRNA and cDNA amplification were performed under
              low stringency conditions."

ORIGIN

Query Match   37.7%; Score 262.4; DB 10; Length 286;
Best Local Similarity   99.6%; Pred. No. 5.1e-42;
Matches   263; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   396 ATGAAAGAAAGGTGAGCACAGCCATTCCTTTACACAGAAATAACAGATTCCCAGCCT   455
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   286 ATGAAAGAAAGGTGAGCACAGCCATTCCTTTACACAGAAATAACAGATTCCCAGCCT   227

Qy   456 CTAACCAAAGAAACAACAAGTTTGGGAACATTCCTCCTTCTGAAATATGAAAGAGAGGG   515
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   226 CTAACCAAAGAAACAACAAGTTTGGGAACATTCCTCCTTCTGAAATATGAAAGAGAGGG   167

Qy   516 GATAAATACTGGAGTAGGATTGTGAAAAAAGTCAAGAGAAAAAAAGACAGCCCAAGT   575
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   166 GATAAATACTGGAGTAGGATTGTGAAAAAAGTCAAGAGAAAAAAAGTACAGCCCAAGT   107

Qy   576 GTAACAGATACTTCTCCATGGGATGGATGGTAAAAAAGGAAGTTATTTCAATAAAATGACCCCTT   635
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   106 GTAACAGATACTTCTCCATGGGATGGATGGTAAAAAAGGAAGTTATTTCAATAAAATGACCCCTT   47

Qy   636 GGAAGGAGTTCAAAACAGGTTGCC   659
           ||||||||||||||||||||||||
Db   46 GGAAGGAGTTCAAAACAGGTTGCC   23

RESULT 2
CNS005TE/c
LOCUS   CNS005TE   997 bp   DNA   linear   GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL060767
VERSION   AL060767.1 GI:4943573
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 997)
  AUTHORS   Genoscope.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a

```
RESULT 3
CNS00418/c
LOCUS       CNS00418      987 bp    DNA    linear    GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR09C16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066537
VERSION     AL066537.1  GI:4942778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 987)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..987
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR09C16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"

ORIGIN

Query Match             8.8%;  Score 61.2;  DB 29;  Length 987;
  Best Local Similarity   22.0%;  Pred. No. 0.027;
  Matches   90;  Conservative  136;  Mismatches  184;  Indels    0;  Gaps    0;

Qy      2 AAAATATCTGATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGAGAGGCTTTGATGG 61
            ||::  |: |||::|||||| :::|||::|||||:|:|||
Db    917 AARARAAAAARAARRAARRAARRAAGARARGRGRAGRAARRGRGRGRGRAGRRR 858

Qy     62 CGATACAGACTTGGAAGACATCAGTCGTGCTGAGCCAGTAAATGAGATGATTCAGGAAGAGTA 121
            :::::|:   ::||::|  :::||::|:|
Db    857 RRRRRGRGRGAGRGRGRRRRGRRGRGRRRRRRGRRGRRRAAARRRRRAGARARRRR 798

Qy    122 TAAACTGGGAAGAGGACAGAGGACAGGCTCAAGGAACATATTTAAGGACTCGGCTAGAAAA 181
            :::::|::  |:::|:|   ||::  ||:|||
Db    797 ARGRGRGRRGARGRGRGGAAAARARARARRRGGRGGARGGAGARGGAGAAARRAAA 738

Qy    182 ACAAGAGAGTATGAACAAGAGTGAGGAGATTATTAGCCAGTGACCCTTTGAGAGTACATCTC 241
            : |||:|:|   :::|::|||||   |:::   ::|||
Db    737 AAARAAAAGARAARAAARAAARAGRARARRGRARGRGRAGRGRGRRGRGRGRGRG 678

Qy    242 TAGGATGGTGTATGTGTAGACCAGATTTTCAGCATCAGAACCATCAGAATTTTGGGGGTG 301
            :::::  ||   :|:  :::||:  :::||:||
Db    677 GRGARRRRGAGRRRRGRGARGRGRGRRRGGAGRGRRRGGAGAGARRRRARGRRAGGAGRG 618

Qy    302 GAAGGAAAAGGAGGCCATGAATCAAAAGGTGGGGAAAAGGCCTTTGGGAGGTGGCAGTGT 361
            |:::|:|   :::::|:|||||:||:||
Db    617 ARGRGRGRRRRARRAGAGARGRGARRRARRGRRAGGARRAAAAAARRGRGGARAGGGAGR 558
```

```
RESULT 4
CNS0039G/c
LOCUS       CNS0039G     1101 bp    DNA    linear    GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"

ORIGIN

Query Match             8.0%;  Score 55.4;  DB 29;  Length 1101;
  Best Local Similarity   16.2%;  Pred. No. 0.37;
  Matches  110;  Conservative  284;  Mismatches  284;  Indels    1;  Gaps    1;

Qy     13 TAGGCAGTTAGAAATTTGAGTTGGAACACAGGAGAGCGTTTGATGGCGCATACAGACT 72
            ::|:: |: :|: |||| :|:::|:|::|
Db   1100 KARRWGDDTWDRDTRKDDWDTKKWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWW 1041

Qy     73 TGGAAGACATCAGTCGTGCTGAGCCAGTAAATGAGATGATTCAGGAAGAGATAAAACTGGGAA 132
            ||:: ::|::  :::|:::::|
Db   1040 WWWWNATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGGAGKRDRDAATDADGA 981

Qy    133 GAGGACAGAGGACAGGCTCAAGGAACATATTTAAGGACTCGGTAGAAAAACAAGAGAGTA 192
            |:::||:|:  ::::::|| :::|||::|
Db    980 GRRDGGRKRRDKKDRKDGDDDKKKGKKKKAKAAKWATKWWDDWDWDKDWKWDGAKDRKA 921

Qy    193 TGAACAAGAGTGAGGAGATTATTAGCCAGTGACCTTTGAGAGTACATCTCTAGGAGTGGTAT 252
            ::::|:::::|:||
Db    920 DDDDGA-GDKDDDGKGKDADDDTDGTKDDDKDWDDKAKGTWGDATWAWAATDWTWWW 862

Qy    253 GTGTAGACCAGATTTTCAGCATCAGAACAACCATCAAGCATTTTGGGGGTGGAAGGAAAAGG 312
            |:::|::||  ::|:::|:|
Db    861 GWADADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRDWDRAGKRGGARKRR 802
```

Query Match   7.9%;   Score 54.8; DB 28; Length 846;
Best Local Similarity 45.8%; Pred. No. 0.51;
Matches 176; Conservative   0; Mismatches 208; Indels   0; Gaps   0;

```
Qy      30 GAGTTTGGAACACAGAGAGACGCTTTGATGCGCGATACAGACTTGAAGACATCAGTGCT  89
           :::     :   ::::  : ::::: :::     :::::::::::::::
Db     787 GAGTTTCCGAGGGAGGAGGAGCACAGAGGGGAGAGAGGGAGAGAGAGACAGGAGCAAG 728

Qy      90 GAGCAGTAAATGAGATGATTCAGGAAGAGAGTATAAACTGGGAAGAGAGACAGGAGGC 149
                       ::  :  ::   ::  :  :::::::::::::  :::: ::::::
Db     727 AGGGAGGAGGGGGGAGGAGGAGGAAGGGAAGAGGGAGGGGAAGAGGGTGAGGGGGAGA 668

Qy     150 TCAAGGAACATATTTAAGGACTGGGTAGAAAAACAAGAGAGTGAGAAAAGTCAAGA  209
                        :: ::    ::       :::  ::: :::  :::::::: ::::
Db     667 AGGAGAGGAGGAGGGGGGAGAGAGAGAAAAGGGGGAAGGGAGGGGAGAGAGAGAGNA 608

Qy     210 ATTATTAGCAGTGACCTTTGAGAGTACATCTTCAGTGGTGGTGTATGTGAGCCCAGATTTT 269
              ::::::::::  : : ::: : ::: ::: :::::::::::: ::::  :::
Db     607 AGGAGAGGGAGGGGGGGAAGGGAGGGGGCGGCCNGGGAAANGGGGGGGGGG 548

Qy     270 CAGCATCAGAACCATCAAGCATTTTGGGGGTGAAGGAAAAGAGCCATGAATCAAAAGG 329
                :::  ::   :::  : :  :::::::::::::::  :::::::
Db     547 CAGGGNGAGGAGAAAAAAGAGAGAGGGGGGGCAAGGNGAAGACGAAGGGGGGAGA 488

Qy     330 TGGGGAAAAGGCCTTTGGGAGTGGGCAGTGTGGGTTAGGGACTTCCCGGTTAGAAT 389
              ::::::::::::::     :::::::::::::::::::
Db     487 GGGGGGAAGGGGAGGGGACGGGGGGGGGGCGAGGAGAGGAAGGGAGGGGGGGGACAAA 428

Qy     390 ACATGGATGAAAGAAAGTGAGCA 413
              ::::::::::::::
Db     427 GAACGGCGAGATCGAACGGGGAGA 404
```

RESULT 6
CNS01637
LOCUS   CNS01637   1373 bp   DNA   linear   GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15I13 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106237
VERSION   AL106237.1 GI:5620820
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1373)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequence;
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES   Location/Qualifiers
     source   1..1373
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15I13"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

Query Match   7.3%;   Score 50.8; DB 29; Length 1373;

```
Qy     313 AGCCATGAATCAAAAGGTGGGGAAAAGGCCTTTTGGGAGGTGGCCAGTGTGGGTAGGGAGT 372
              :: :::   :  ::  ::::::     :::::::
Db     801 DRKRADDKRDAADDRDDAAIMTTWTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRD 742

Qy     373 AGGGGCTCCGGTTAGAATACATGGATGAAAGGTGAAAAAAGTGAGCACAGCCATTTCCTTACAAC 432
                :    ::::::   ::::  ::: :::::::  ::::::  : :::::
.Db    741 RAGTAGRKNRRTWKRRWKRRDTRWDDADADDTARDDRRRGDGDADAGKGKKTGRKRRRR 682

Qy     433 AGAAATAACAGATTTCCCAGCCTCTAACCAAAGAAACACACAAGTTTGGGAACATTCCTCT 492
              :: ::   : :      ::  :  :  :: :::: :   ::   ::::
Db     681 DRATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWDW 622

Qy     493 CTTCTGAAATATGAAAGAGAGAGGGGATAAAATACTGGAGTAGGATTGTGAAAAAGTCAAGA 552
              :: :::   :    :  ::::  :  : ::  ::  ::::  :
Db     621 AKWDWKTRADRWDRWRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRRWTTKGKTTT 562

Qy     553 GAAAAAAAGAACAGCCCAAGTGTAACAGACATACTTCTCCATGGGATGGTAAAAAAGGAAG 612
              : ::::::::::::  ::      :::  :::  ::  :  :::
Db     561 ATWTTWAARAARAAWWAWWAWATTTATWTTWTTWTTTTTTTTTTTTTTTAAWWWAAAWTAT 502

Qy     613 TTATTTCAATAAATGACCCCTTGGAAGGAGTTCAAAACAGGTTGCCATATGCTTTATGT 672
              :: : ::     ::   : :::   :::  :::: :::::::
Db     501 WAAWTAAAWAAAAAAAWAAATTTTTTTTTTTTTTWAWTAAAWTAWTWTTTTWTTTTWWAATT 442

Qy     673 GGAGTTTTGAAAATATAT 691
              :::  :::::::::::
Db     441 TTTTWTWTTWWAATTATTT 423
```

RESULT 5
B18548/c
LOCUS   B18548   846 bp   DNA   linear   GSS 09-AUG-1997
DEFINITION T408-Sp6 TAMU Arabidopsis thaliana genomic clone T408, genomic
            survey sequence.
ACCESSION   B18548
VERSION   B18548.1 GI:2316452
KEYWORDS   GSS.
SOURCE   Arabidopsis thaliana (thale cress)
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 846)
AUTHORS   Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
TITLE   BAC End Sequences at ATGC
JOURNAL   Unpublished (1997)
COMMENT   Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: Sp6
            Class: BAC ends
            High quality sequence start: 116
            High quality sequence stop: 136.
FEATURES   Location/Qualifiers
     source   1..846
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="T408"
                     /sex="hermaphrodite"
                     /clone_lib="TAMU"
                     /note="Vector: BeloBACII; Site_1: HindIII; Site_2:
                     HindIII; Produced by Rod Wing"
ORIGIN
```

(Left column — RESULT 7)

```
Best Local Similarity  30.1%;  Pred. No.2.9;
Matches 119;  Conservative 84;  Mismatches 193;  Indels  0;  Gaps  0;

Qy    16  GCAGTTAGAAATTTGAGTTTGGAACACAGGAGAGAGCTTTGATGGGCGATACAGACTTGG  75
              |||||  |  |||||||  ::||  |||| ||| |||||  : |
Db   915  GSAGGNDTGKRAKKGAGGGKGRGKGRGKRAKGRGGAATTTKKKRAGAGGGGARAGAGRARKWG  974

Qy    76  AAGACATCAGTGCTGAGCAGTAGGCAGTAAATGATGAATTCAGGAAAGAGTATAAACTGGGAAGAG  135
              ::|||  |  :  | | ::::|
Db   975  RAGATWIKKTTTTTTTTKKGAGGARARGGAGCGAGGRRGAAAGAWGRARAGAGGGGARRGRA  1034

Qy   136  GACAGAGGACAGGCTCAAGGAACCATATTTAAGGACTCGGTAGAAAAACAAGAGAGTATGA  195
              ||||| : | | ::|
Db  1035  GATRGRRKRTAGGGGARAGGGGRARGGAGAGRGRGGGGGRGRGRRGRGGGAAGAGARGGR  1094

Qy   196  ACAAGAGTGAGGAGATTATTAGCAGTGACCTTTGAGAGTACATCTCTCGAGAGTGGTATGTG  255
              :::||||  | :||| ||
Db  1095  GGWGRAGRGGGRRAGWRKAGKARAAAARAATTGKGAGGGRGRGGARRGCANAAAAAGAGAG  1154

Qy   256  TAGAGCCAGATTTCAGCATCAGAACCATCAAGCATTTTGGGGGTGGAAGGAAAAGGAGAGC  315
              |  |||  ||
Db  1155  ARGAGRGAGGGGGWRRTGAGDAGARAGAGAGAGARARADDRRARRGAGAGAGAGAGA  1214

Qy   316  CATGGAATCAAAAGGTGGGGAAAAGGCCTTTGGGAGGTGGCAGTGTGGGTAGGGAGTAGG  375
              |  : :: ||| | ||  | | |  || |||||
Db  1215  GAGRRRRRGAGDRDARAGATAGAGATATWGRWGDAAAWARAGTKWGDKAKRGGRGRGAGRA  1274

Qy   376  GCTCCGGTTAGAATACATGGATGAAAGAAAGGTGAG  411
              |  ||:| |||  |||:::||||
Db  1275  TAGADAGARWAAAAAKDAGWGTGAGAGWRWGKDGAG  1310
```

```
RESULT 7
CNS000D1
LOCUS        CNS000D1      1101 bp   DNA     linear   GSS  03-JUN-1999
DEFINITION   Drosophila melanogaster genome survey sequence TET3 end of BAC #
             BACR01J16 of RPCI-98 library from Drosophila melanogaster (fruit
             fly), genomic survey sequence.
ACCESSION    AL065414
VERSION      AL065414.1  GI:4938827
KEYWORDS     GSS.
SOURCE       Drosophila melanogaster (fruit fly)
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
             Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
             Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 1101)
AUTHORS      Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
             - Web : www.genoscope.cns.fr)
COMMENT      Determination of this BAC-end sequence was carried out as part of a
             collaboration with the Berkeley Drosophila Genome Project (BDGP).
             The BDGP is constructing a physical map of the Drosophila
             melanogaster genome using these BACs. For further information
             please see http://www.fruitfly.org the BDGP Drosophila
             melanogaster BAC library was prepared by Kazutoyo Osoegawa and
             Aaron Mammoser in Pieter de Jong's laboratory in the Department of
             Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
             NY. The library is named RPCI-98 and was constructed by partial
             EcoRI digestion of Drosophila DNA provided by the BDGP from the
             isogenic strain y2; cn bw sp; the same strain used for the BDGP's
             P1 and EST libraries. A more detailed description of the library
             and how to order individual BAC clones, the entire library, or
             filters for hybridization from the BACPAC Resource Center can be
             found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR01J16"
```

(Right column)

```
                     /clone_lib="RPCI-98"
                     /note="end : TET3"

ORIGIN

Query Match      7.2%;  Score 49.8;  DB 29;  Length 1101;
Best Local Similarity 16.5%;  Pred. No.4.7;  Indels  2;  Gaps  1;
Matches  90;  Conservative 227;  Mismatches 226;

Qy     1  TAAAATATCTGATAGGCAGTTAGAATTTGAGTTTGGAACACAGAGAGAGGCTTTGATG  60
              |||||  |  ::|||  |  |||||||  ||| ||| |||  :  : :::
Db   557  TAAWYTTTTTTWTATTTTWKAANAAAAAAAAAAAAAAAWAGARGATNDDKBAKDWKA  616

Qy    61  GCGATACAGACTTGGAAGACATCAGTGCTGAGCAGTAGGCAGTAAATGATGAATTCAGGAAAGAGT  120
              ::||||  |  :::||| | | | ::
Db   617  AADAKAKAAAWHAGTWWTKKDTKKKKKKKRKWGKAKRAATDDTAWTATTWKKTAKKDAGK  676

Qy   121  ATAAACTGGGAAGAGGACAGAGGACACAGGCTCAAGGAACATATTTAAGGACTGGGTAGAAA  180
              |  : | | | | :| |||  :  |:| :| :|
Db   677  AAKWKADRTKAWKGAKGAGARGKATAKARAKKRWKDDKDTKKKKKTTATKKKTTTKRA  736

Qy   181  AACAAGAGTATGAACAAG--AGTGAGGAGATTATTAGCAGTGACCTTTGAGAGTACAT  238
              |  || | |  |  |:: :: ||::::||||
Db   737  ATDKWKWRKKKATAGAWDWDDARRGRGGRWGDKRRRTGARGDKDKTGKKKKTGDKRDKGG  796

Qy   239  CTCTAGAGTGGTATGTGTGTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGG  298
              :|| ::|:|  :::|  :  | |  |||  ::||
Db   797  GDGDRDGDGAGGGDRDRDARARRARWDTKDTRGAKDKADKKKAGRKRKWWTRKTWKRDTA  856

Qy   299  GTGGAAGGAAAAGGAGCCATGAATCAAAAGGTGGGGAAAAGGCCTTTGGGAGGTGGCAG  358
              ::||  :  |||:|  ||  ||||| : ||  :::|
Db   857  DWKAWAAAWRWRWRDRKDRDRGRRDRWRRWRGADKKWKWDDDTAKDWTKKDTDWTAGAT  916

Qy   359  TGTGGGTAGGAGTAGGGCCTCCGGTTAGAATAACATGGATGAAAGAAAGGTGAGCACAGCC  418
              ||  ||| | | | ||  || |  :|| : : :::|  |  ||| ::
Db   917  AGWDDDKAKDGDWKRGTDWRTWKRTWKGWGRKGDKGRKTAGRAGDGKDKKKGKDKDRTDD  976

Qy   419  ATTTCCTTACAACAGAAATAACAGATTTCCCAGCCTCTAACCAAAGAAACAACAAGTT  478
              ||| | | | |:  |:  ||  :  |  ::::|:
Db   977  AKATDGKDTWRDTTDTAKRRRKRRRRAGRGRDRKGDKKKRWRARWRKGKKKKTKDKDD  1036

Qy   479  GGGAACATTCCTCTCTTCTCTGAAATATGAAAGAACGAGGGGATAAAATACTGGAGTAGGATTGT  538
              |||||| | |  ||  ||| || |  |  ::  || : |
Db  1037  DGKRWGRAKAKADAAKAKDDADAADAAAWAKAKAKDKDKAKKKADADDRKDKKAARKDD  1096

Qy   539  GAAAA  543
              :||
Db  1097  KAKAA  1101
```

```
RESULT 8
CNS017RP
LOCUS        CNS017RP      1101 bp   DNA     linear   GSS  26-JUL-1999
DEFINITION   Drosophila melanogaster genome survey sequence SP6 end of BAC
             BACN37J10 of DrosBAC library from Drosophila melanogaster (fruit
             fly), genomic survey sequence.
ACCESSION    AL108415
VERSION      AL108415.1  GI:5628719
KEYWORDS     GSS.
SOURCE       Drosophila melanogaster (fruit fly)
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
             Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
             Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 1101)
AUTHORS      Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
             - Web : www.genoscope.cns.fr)
COMMENT      Determination of this BAC-end sequence was carried out as part of a
             collaboration with the European Drosophila Genome Project (EDGP) -
             http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
             library (Dros BAC) was made by Alain Billaud at CEPH (Centre
```

```
FEATURES
     source          d'Etude du Polymorphisme Humain) with funding provided by a MRC
                     project grant. The DNA was prepared from embryos by Alain Bucheton
                     and Genevieve Payan. It has been constructed in the vector
                     pBeloBAC11.
                     Location/Qualifiers
                     1..1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN37P10"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"

ORIGIN
Query Match          7.1%;  Score 49.4;  DB 29;  Length 1101;
Best Local Similarity   25.1%;  Pred. No. 5.7;
Matches   89;  Conservative 100;  Mismatches 166;  Indels   0;  Gaps   0;

Qy    22 AGAAATTGAGTTGGAACACAGGAGGAGGAGGCTTTGATGGCGATACAGACTTGGAAGACA 81
            | ::| ||||   |: :| | | :| |::||| | | || ||||
Db   742 ARARARGSGRGARRRAARARGARARARRAARRGRRGRARARAAGGAGAARRRRRAGRRGG 801

Qy    82 TCAGTGCTGCTGAGCAGTAAATGAGATGATTCAGGAGGAAAGAGTATAAACTGGGAAGAGGACAGA 141
          :: |  | ::| |: | ||||| :| | | ||:: ||||| :: ::
Db   802 GRRGRGRGRGRGGGARAGAGAAGACGRGGARAGARRRARARARAAGARARARAAARRG 861

Qy   142 GGACAGGCTCAAGGAACATATTTAAGGACTGGGTAGAAAAAACAAGAGAGTATGAACAAGA 201
          :: |:: | |||||:|| ::|||   | ||||| | ||||| || :|
Db   862 RRGRAARARAGAGAARRRGAAGAAARRARAGRAGAARARARAAGARARRRARRRGR 921

Qy   202 GTGAGGAGATTATTAGCAGTGACCTTTGAGATACATCTCTAGAGTGGTATGTCTGAGAGC 261
          ::|||||||   | :|| :| :  || |: | | :||: | ::|| : |:|
Db   922 RGRRRRAGAGARGRGRGRGGGARRGRGRGRRRRRRGGGARRGGRGGRGG 981

Qy   262 CAGATTTCAGCATCAGAACCATCAAGCATTTTGGGGTGGAAGGAGGAGGCCATGAA 321
          :: :|||:|| | : : :|   ||||||||||||||| :|||:||:|| :|
Db   982 RRRGRARRGRGRGRRRGGRGRGGGGGGGGGGGAGRGGGGGGAGRRGGRGRAG 1041

Qy   322 TCAAAGGTGGGGAAAAGGCCTTTTGGGGAGGTGGCAGTGGGTAGGGAGTAGG 376
          ::: :| |:::|| : :| |  ::|||||||||||||| ::::|| |||| :||
Db  1042 RARGGGRGRRRGGRRRGGGRGRGGRGRGRGGRGGRGRGGRGGGGGGGAGRRGGGGGGGARG 1096

RESULT 9
CNS0181N
LOCUS       CNS0181N                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37P10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108773
VERSION     AL108773.1  GI:5629077
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
```

```
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN37P10"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"

ORIGIN
Query Match          7.0%;  Score 48.6;  DB 29;  Length 1101;
Best Local Similarity   22.2%;  Pred. No. 8.2;
Matches   70;  Conservative 105;  Mismatches 140;  Indels   0;  Gaps   0;

Qy   102 AGATGATTCAGGAAGAGTATAAACTGGAAGAGGACAGAGGACAGGCTCAAGGAACATA 161
          ::: :|:  ::|: :|: :| :: |::: : : :::: :|:  :|: :| |
Db   774 RVRARAGRGRRRGGDRARRARHRRRRARAAAARRGRRMMVRNMRARNAAA 833

Qy   162 TTTAAGGACTGGTAGAAAAACAAGAGAGTATGAACAAGAGATTATTAGCAGT 221
          ::: ::|: |:| |: ||| :|::| : |:|::| ||| :|: :|||
Db   834 RAGRARGRRAGRRARARRAGAARRGRGRRRGRRRRRARARARAARRARRG 893

Qy   222 GACCTTTGAGAGTACATCTCTAGAGTGGTGTATGTGTAGAGCCAGATTTTCAGCATCAGAAC 281
          ::  |::|| | || |:||||| :| |||||::|| |::::|
Db   894 RRAAAARRAGGRGRAGRRARRAARGGRRRRRGRRRRGRRRGRAGGGAGAGGRRA 953

Qy   282 CATCAAGCATTTTGGGGTGAAGGAAAAAGGAGCCATGAATCAAAAGGTGGGAAAAGGC 341
          ::: |  :::::||||| :::|| :||| :::|| |:|||| :||||||
Db   954 GAGARRGRGRGGGGGGGGGGGGGRGRRGRRRAARRRGRAARRAAGGARARARARAARARA 1013

Qy   342 CTTTTGGGAGGTCGCAGTGTGGGTAGGGAGTAGGGCTGTGTTAGATACATGGATGAAA 401
          ::: :||||| :|: ||:|| :|:::| ||: |||| :::|
Db  1014 RAGAGARARARRGRAGARGRAGGGGGGGGGRGGGGGGGGRRGGGGGRRGRRGRRGGRGG 1073

Qy   402 GAAAGGTGAGCACAG 416
          ::| :|| ||||| :
Db  1074 ARRGGGRRGGARRRR 1088

RESULT 10
CE833588/c
LOCUS       CE833588                 687 bp    DNA     linear   GSS 30-SEP-2003
DEFINITION  tigr-gss-dog-1700033229120 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE833588
VERSION     CE833588.1  GI:37174608
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
  MEDLINE   22875432
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirkness@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..687
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
```

```
459 ACCAAAGAAACAACAAGTTTGGGAACATTCCTCTCTCTGAAATATGAAAGAGAGGGGGAT 518
        :||: |:| :||| : :||:   ::  :.::::  : :.:  : :::
107 AAAAWWAWAWARGAWAWGCWWGGAWWMMWARGMMMWMWAWWMAAAAAAAWWMWCTTTGCMCT 166

519 AAATACTGGAGTAGGATTGTGAAAAAAGTCAAGAGAAAAAAAAGAACAGCCCAAGTGTA 578
        ::|:  :| ::::  ::|||:::.:||::| ::||  :::.::||:::
167 GAWTRTWGTTRTWTWAWWMRTTRAAWWNAAWWMWAAWWWWARGWWWWRGMAAAAARAW 226

579 ACAGATACTTCTCCATGGGATGGTAAAAGGAAGTTATTCAATAAAATGACCCCCTTGGA 638
        |:::::|  ||||||  :::  :   ::::.:  |||||
227 WWAMWRCRCWTWCRCMAAMMAAVAAAAAMAAMAAAMARCAWWMWMAAWCRAAWMGGA 286

639 AGGAGTTCAAAACAGTTGCCATATGCTTTATGTGGAGTTTTGAAAAATATATAAAA 695
        :  :::: :::::  :||:|::|| |:  :  : :|:|::|:|||
287 WARAAWTRGAWWWMRAMWAAAMRAARMAWRRAWWMAAAWMAAWTTACAWAAAAWAAAA 343
```

RESULT 12
BM414951/c
LOCUS   BM414951   1478 bp   mRNA   linear   EST 28-JAN-2002
DEFINITION OP20019 Mixed Stage EST's from Globodera pallida, the potato cyst
            nematode Globodera pallida cDNA, mRNA sequence.
ACCESSION   BM414951
VERSION   BM414951.1 GI:18381182
KEYWORDS   EST.
SOURCE   Globodera pallida
  ORGANISM Globodera pallida
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
REFERENCE   1 (bases 1 to 1478)
  AUTHORS   Heer,J., Sosinski,B., Pokrzywa,R.M., Warry,A. and Opperman,C.
  TITLE   Mixed Stage EST's from Globodera pallida, the potato cyst nematode
  JOURNAL   Unpublished (2001)
COMMENT   Contact: Opperman, C
            Center for the Biology of Nematode Parasitism
            NC State University; IACR-Rothamsted
            Campus Box 7616; Raleigh, NC 27695, USA
            Tel: 919.515.6699
            Fax: 919.515.9500
            Email: warthog@unity.ncsu.edu
            GT11-10PCN_F_B08_GT11-10_F_061.ab1.
FEATURES   Location/Qualifiers
     source   1..1478
                     /organism="Globodera pallida"
                     /mol_type="mRNA"
                     /db_xref="taxon:36090"
                     /clone_lib="Mixed Stage EST's from Globodera pallida, the
                     potato cyst nematode"
                     /note="Vector: lambda GT11; This is a collaborative effort
                     between IACR-Rothamsted and North Carolina State
                     University. The library was constructed from mixed stage
                     G. pallida in lambda GT11 by Paul Burroughs,
                     IACR-Rothamsted."
ORIGIN
     Query Match   6.9%; Score 47.8; DB 12; Length 1478;
     Best Local Similarity 40.4%; Pred. No. 11; Indels 0; Gaps 0;
     Matches 133; Conservative 0; Mismatches 196;

```
116 AGAGTATAAACTGGGAAGAGGACAGCAGGCTCAAGGAACACATATTTAAGGACTGGGT 175
        |   |||||||||||||||  |   :||||||| ||||||||||||||||||||
760 ANAGAAGAAAAANAGGGAGGAGAGGAGAGAGAAAAAAAGAAAAAAGGGAGNGGGGG 701

176 AGAAAAACAAGAGAGTATGAACAAGAGTGAGGAAGAAGAAAAAAAAAAAAAAGGGNNGCCNNNNNNN 235
        ||||||||||||||||  |||||||||||| ||||||||||||||||||||||||
700 AANNAARAGAGANGANGNGAAGGAAGGAAGAAGAAAAAAAAAAAAAAAAAGGGNNGCCNNNNNNN 641

236 CATCTTCTAGAGTGTATGTGTAGAGCCAGATTTCAGCAGCATCAGAGAACCATCAAGCATTTG 295
        |||| ||| |||||||||||||||  |||| |||||||||||||||  |||
640 CCNNNNNAAAAAANNNNTTTTTTGAGGGATTTTTTTTTAAAAGGNGAGAAGAAGAG 581

296 GGGGTGGAGGAAGAAAAAGGAGGCCATGAATCAAAAGTGGGGAAAAGGCCTTTTGGGGAGTGG 355
        |||||||||||||||||||||||||||||||||
```

RESULT 11
CNS0187R
LOCUS   CNS0187R   878 bp   DNA   linear   GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN04E04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108993
VERSION   AL108993.1 GI:5629297
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
  ORGANISM Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 878)
  AUTHORS   Genoscope.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES   Location/Qualifiers
     source   1..878
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN04E04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN
     Query Match   6.9%; Score 48.2; DB 29; Length 878;
     Best Local Similarity 25.9%; Pred. No. 10; Indels 0; Gaps 0;
     Matches 77; Conservative 86; Mismatches 134;

```
399 AAAGAAAAGTGAGCACAGCCATTTCCTTTACAACAGAAATAACAGATTTCCCAGCCTCTA 458
        ::||| |: :| :|| : ::|:||:: :::
47 MAATCAAAWRTATGAWAAAAWMMCAAAAAAAAWMMAAAAAWWMMAAAAWWAWAAAAAWAWMWMAAA 106
```

```
            /clone_lib="Dog Library"
            /note="Site 1: BstXI; Libraries were prepared from
            peripheral blood"
ORIGIN
     Query Match     6.9%;  Score 48.2;  DB 29;  Length 687;
     Best Local Similarity 52.8%;  Pred. No. 11;  Indels 0; Gaps 0;
     Matches 104;  Conservative 0;  Mismatches 93;

433 AGAAATAACAGATTTCCAGCCTCTAACCAAAGAAACAACAAGTTTGGGACATTCCTCT 492
        |||||||||||||||||||  |  ||||||||||||||
655 AAACAGGACAGCTTTCTGTTCCCCTGACCAAGATAAAAGGAATGCTCGAGAGGTAATCTT 596

493 CTTCTGAAATATGAAAGAGAGGGGGATAAATACTGGAGTAGGATTGTGAAAAAGTCAAGA 552
        |||||  |||||||||||  :||||||||||||
595 CTGATGAAGAGAAAATTAAGGGTCTTACTAATAGACCAGGTATGAGAAGAGAAAGA 536

553 GAAAAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTCCATGGGATGGTAAAAGGAAG 612
        |||||  ||| ||||||||  |||||||||
535 AAGAAAAGAGAAAAAGCCAAGGATGATGATTCTAATATTTTGACCTGAGCAACTGGAAGAAT 476

613 TTATTCAATAAAATGA 629
        ||||||||||||
475 GGAATTCCATTTACTGA 459
```

```
Db   580  GGGGGGGAGGAAANANNGGCAAGAAGGAAGAGGAAGGGNGAGGAGGGGG  521
                                                        
Db   308  AAWAATWGATTAAWGWTTAWAWWWAAATTAWAWWWTTGAAAAAAAAATTGTGAAAAAA  249

Qy   356  CAGTGTGGGTAGGGAGTAGGGCTCCGGTTAGAATACATGGAAGAAGGTGAGCACA  415
Qy   624  AAATGACCCCTTGGAAGGAGGTTCAAAACAGGTTGCCCATATGATGCTTTATGTGGAGTTTTGAA  683

Db   520  GGGGAAGGAGGNGNGGGGAGGGNANGAGGAGNAANNGNNAGAAAAAGAGNGNNNAGA  461
Db   248  AWATWWAAAATWTAATAATTTAAAATCTTTCCAAMTTTTCAWAAWTTTWATRAAAAAA  189

Qy   416  GCCATTTCCTTACAACAGAAATAACAGA  444
Qy   684  AAATATATAAAA  695

Db   460  GGNANAGAAGANGNNAGNNAAAAACAAAAGA  432
Db   188  AAMTTWMAAAAA  177
```

RESULT 13
AL514901/c
LOCUS   AL514901   932 bp   mRNA   linear   EST 08-MAY-2003
DEFINITION   AL514901 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
             CL0BB012ZH12 3-PRIME, mRNA sequence.
ACCESSION   AL514901
VERSION   AL514901.2   GI:30464786
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 932)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE   Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT   On Feb 13, 2001 this sequence version replaced gi:12778394.
             Contact: Genoscope
             Genoscope - Centre National de Sequencage
             BP 191 91006 EVRY cedex - France
             Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr
             Library was constructed by Life Technologies, a division of
             Invitrogen. This sequence belongs to sequence cluster 4052.r For
             more information about this cluster, see
             http://www.genoscope.cns.fr/
             cgi-bin/cluster.cgi?seq=CL0BB012ZH12FP1&cluster=4052.r. Contact :
             Feng Liang Email : fliang@lifetech.com URL :
             http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
             Faraday Avenue Genoscope sequence ID : CL0BB012ZH12FP1.
FEATURES   Location/Qualifiers
     source   1..932
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CL0BB012ZH12"
                     /tissue_type=NEUROBLASTOMA"
                     /clone_lib="Homo sapiens NEUROBLASTOMA"
                     /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

Query Match   6.8%; Score 47.6; DB 9; Length 932;
Best Local Similarity 36.2%; Pred. No. 13;
Matches 113; Conservative 45; Mismatches 154; Indels 0; Gaps 0;

```
Qy   384  TAGAATACATGGATGAAAGAAAGGTGAGCACAGCCATTTCCTTTACAACAGAAATAACAG  443
               |||| ||||| | ||||||||| |  | |::||||| | |   ||| :  ||  | 
Db   488  WAAAATAAAAWAAARRAAAWAMAWTAAAAMAWTAAAAMAAAAAAAAAAGWRCWAAAAATAARMWWWT  429

Qy   444  ATTTCCCAGCCTCTAACCAAAGAAACAAAGAAACAAGTTTGGGAACATTCCTCTCTTCTGAAATA  503
               |||| |: :::  |     |  ::|||||| | | |  |  :: ::
Db   428  AAGGACCCTTTAMMAAAAAWAAACTGGAGTAGGATTGTGAAAAAAWAAATCWRRTATWAAWMTTWTMAAAAWAAAA  369

Qy   504  TGAAAGAGAGGGGATAAATACTGGAGTAGGATTGTGAAAAAAGTCAAGAGAAAAAAAAG  563
              |||||| |:: ||| |||          |:|       |||||||:|
Db   368  RAAAWAAAAAAAMWWTWWARAAWAAATAARRAAAAAAAAAAATTAAAAAAAAAWAAWAM  309

Qy   564  AACAGCCCAAGTGTAACAGATACTTCTCCATGGGATGGTAAAAAGGAAGTTATTTCAATA  623
               |||     |  |   : :   | | |     |||  | || |||   |
```

```
Db   581 WAAATTKKTKKTKKTWKTKKKTTTTRDTWDAARARRRRRRRRRRRRRGGRRRR 522

Db   1071 AAAADKKKKKKAAAAAAAADKDDDAAAAAAKKKKKKKKKKDAAAAAAADDKDKKKKKKKDD 1130

Qy   301 GGAAGGAAAAGGAGCCATGAATCAAAAGGTGGGAAAAGGCCTTTTGGGAGGTGGCAGTG 360
         :::::::::::::: : ::: |||: : ::::: : |: :: :: :||:
Db   521 RRRRRRRRRRRRRRRRRRAARRAANNRRRDRNKRNNTKTKWAWTWATRKRGRTK 462

Qy   600 GGTAAAAAGGAAGTTATTCAATAAAA 626
         :::||||| ||| :| : ||||
Db   1131 KKKAAAAAAAAAAAAKAAAKKKAAA 1157

Qy   361 TGGGTAGGGAGTAGGGCTCCGGTTAGAATACATGATGAAAGAAGGTGAGCACA 415
         : |::: |||::|| :||:|||
Db   461 TGTTTTTTTTDTDKDTTGGGKGTGRAAGAAAAAGGRRKDAAMGRGGRGKGAGGAWA 407

         Search completed: April 26, 2004, 01:59:16
         Job time : 1266.83 secs

RESULT 15
BX461310
LOCUS       BX461310                 1201 bp    mRNA       linear   EST 22-MAY-2003
DEFINITION  BX461310 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF025YA16 5-PRIME, mRNA sequence.
ACCESSION   BX461310
VERSION     BX461310.1  GI:31031291
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
AUTHORS     Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr; Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF025BA08QP1.
FEATURES    Location/Qualifiers
     source  1..1201
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="CS0DF025YA16"
             /tissue_type="FETAL BRAIN"
             /dev_stage="fetal"
             /clone_lib="Homo sapiens FETAL BRAIN"
             /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
             was primed with a NotI-oligo(dt) primer. Five prime end
             enriched, double-strand cDNA was digested with Not I and
             cloned into the Not I and EcoRV sites of the pCMVSPORT 6
             vector. Library was not normalized."
ORIGIN

Query Match         6.8%;  Score 47.4;  DB 13;  Length 1201;
Best Local Similarity   22.6%;  Pred. No.14;
Matches   74;  Conservative 117;  Mismatches 136;  Indels   0;  Gaps   0;

Qy   300 TGGAAGGAAAAGGAGCCATGAATCAAAAGGTGGGGAAAAGGCCTTTTGGGAGGTGGCAGT 359
         ::::: :|::||| : :||||| ::||| ||||::::::::::::::::: :: ::
Db   831 KDKDKDARADDGGGGGKKKKKKKDDAAAAADDKKKKKKKKKKKKKRAAAAWTTKKKKK 890

Qy   360 GTGGGTAGGGAGTAGGGCTCCGGTTAGAATACATGGATGAAAGAAAGGTGAGCACAGCCA 419
         ::::::::: :: :::: |||||||||| ::::: |||||| :
Db   891 KKKKKDKKKKKKKKDAAAAAAAAAAAADKKDDAAAAAAAAAAAAKKKKKKKK 950

Qy   420 TTTCCTTTACAACAGAAATAACAGATTTCCCAGCCTCTAACCAAAGAAACACAAGTTTG 479
         ::: |||:: |:|| : ||| :::|:::
Db   951 KKKKKDAAAAAAAADKKKKKKKKKKKKDKKKKKDAAAAAAAAAAAAAAAAAADDKKK 1010

Qy   480 GGAACATTCCTCTCTTCTGAAATATGAAAGAAGAGAGGGGATAAATACTGAGTAGGATTGTG 539
         :: :: :::::|||||||||||KKDAADKKKKKKKAAAAAAKKDAAAADDKKKKKKKA 1070
Db   1011 KKKKKKKKKKKKKKKKKKKKKKDAADKKKKKKKAAAAAAAKKKDAAAADDKKKKKKKDA 1070

Qy   540 AAAAAGTCAAGAGAGAAAAAGAACAGCCCAAGTGTAACAGTACTTCTCCATGGGAT 599
         ||||||: :: | |||||::: :::|| |||| ::::::
```

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:43:48 ; Search time 203.468 Seconds
                                     (without alignments)
                        15422.560 Million cell updates/sec

Title: US-10-051-835-14
Perfect score: 696
Sequence: 1 taaaatatctgatggcagt.........ttttgaaaaatatataaat 696

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :   Published Applications NA:*
                1; /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                2; /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                3; /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                4; /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                5; /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                6; /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                7; /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                8; /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
                9; /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
               10; /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
               11; /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
               12; /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
               13; /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
               14; /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
               15; /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
               16; /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
               17; /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
               18; /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
               19; /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		696	100.0	696	15	US-10-051-835-14	Sequence 14, Appl
2		46.4	6.7	461	10	US-09-814-353-17724	Sequence 17724, A
3		44.8	6.4	9415	9	US-10-311-455-268	Sequence 268, App
4	c	41.8	6.0	300	5	US-09-864-761-26934	Sequence 26934, A
5	c	41	5.9	508	10	US-09-814-353-18511	Sequence 18511, A
6	c	40.6	5.8	6668	15	US-10-311-455-1669	Sequence 1669, Ap
7		40.2	5.8	671	15	US-10-184-644-346	Sequence 346, App
8	c	40.2	5.8	671	15	US-10-184-634-346	Sequence 346, App
9	c	40.2	5.8	1269	13	US-10-282-122A-16543	Sequence 16543, A
10	c	40	5.7	767	13	US-10-027-632-153507	Sequence 153507,
11		40	5.7	767	16	US-10-311-455-153507	Sequence 153507,
12	c	39.6	5.7	6319	15	US-09-854-867-362	Sequence 362, App
13	c	39.4	5.7	2750	13	US-09-854-867-362	Sequence 362, App
14	c	39.4	5.7	6681	15	US-10-311-455-128	Sequence 128, App

c	15	39.4	5.7	6681	15	US-10-240-452-4	Sequence 4, Appli
c	16	39	5.6	583	16	US-10-027-632-136545	Sequence 136545,
c	17	39	5.6	583	13	US-10-027-632-136545	Sequence 136545,
	18	38.8	5.6	921	16	US-10-027-632-120814	Sequence 120814,
	19	38.8	5.6	921	13	US-10-027-632-120814	Sequence 120814,
c	20	38.8	5.6	10048	15	US-10-221-613-212	Sequence 212, App
c	21	38.8	5.6	3573778	15	US-10-312-841-1	Sequence 1, Appli
	22	38.4	5.5	421	15	US-10-060-036-3868	Sequence 3868, Ap
	23	38.4	5.5	657	13	US-10-276-774-703	Sequence 703, App
	24	38.4	5.5	699	10	US-10-764-891-185	Sequence 185, App
	25	38.4	5.5	718	13	US-10-027-632-12438	Sequence 12438, A
c	26	38.4	5.5	718	16	US-10-027-632-12438	Sequence 12438, A
c	27	38.4	5.5	1074	10	US-09-814-353-20743	Sequence 20743, A
c	28	38.4	5.5	1686	15	US-10-302-172-327	Sequence 327, App
c	29	38.4	5.5	2054	15	US-10-198-846-11116	Sequence 11116, A
c	30	38.4	5.5	14861	13	US-10-221-613-161	Sequence 161, App
c	31	38.4	5.5	14861	15	US-10-311-455-167	Sequence 167, Ap
c	32	38.2	5.5	610	13	US-10-027-632-261595	Sequence 261595,
c	33	38.2	5.5	610	16	US-10-027-632-261595	Sequence 261595,
c	34	38	5.5	5823	15	US-10-240-453-256	Sequence 256, App
c	35	37.8	5.4	5660	15	US-10-311-455-2330	Sequence 2330, Ap
	36	37.6	5.4	6136	17	US-10-257-166-124	Sequence 124, App
c	37	37.6	5.4	9539	15	US-10-239-676-51	Sequence 51, Appl
c	38	37.6	5.4	9539	15	US-10-240-453-53	Sequence 53, Appl
c	39	37.6	5.4	14798	15	US-10-311-455-1006	Sequence 1006, Ap
c	40	37.6	5.4	96593	12	US-10-052-482-67	Sequence 67, Appl
c	41	37.4	5.4	410	15	US-10-060-036-3601	Sequence 3601, Ap
	42	37.4	5.4	505	10	US-09-918-995-20057	Sequence 20057, A
c	43	37.4	5.4	2000	16	US-10-260-238-2483	Sequence 2483, Ap
c	44	37.4	5.4	7058	17	US-10-257-166-134	Sequence 134, App
c	45	37.4	5.4	370469	13	US-10-087-192-250	Sequence 250, App

ALIGNMENTS

RESULT 1
US-10-051-835-14
; Sequence 14, Application US/10051835
; Publication No. US20030165864A1
; GENERAL INFORMATION:
; APPLICANT: Jones, David A.
; APPLICANT: Lasek, Amy W.
; TITLE OF INVENTION: GENES REGULATED BY DNA METHYLATION IN TUMOR CELLS
; FILE REFERENCE: PA-0044 US
; CURRENT APPLICATION NUMBER: US/10/051,835
; CURRENT FILING DATE: 2002-01-16
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PERL Program
; SEQ ID NO 14
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030165864A1 213413.1
US-10-051-835-14

  Query Match   100.0%; Score 696; DB 15; Length 696;
  Best Local Similarity 100.0%; Pred. No. 7.3e-187;
  Matches 696; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 TAAAATATCTGATGGCAGTTAGAAATTTGAGTTTTGGAACACAGGAGAGAGGGCTTTGATG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 TAAAATATCTGATGGCAGTTAGAAATTTGAGTTTTGGAACACAGGAGAGAGGGCTTTGATG 60

Qy   61 GCGATACAGACTTGAAGACATCAGTGCTGAGCAGTAAATGATTCAGGAAAGAGT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCGATACAGACTTGAAGACATCAGTGCTGAGCAGTAAATGATTCAGGAAAGAGT 120

Qy   121 ATAAACTGGGAAGAGGACGACAGGCAGAGGACGGCTCAAGGAACATATTTAAGGACTGGGTAGAAA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ATAAACTGGGAAGAGGACGACAGGCAGAGGACGGCTCAAGGAACATATTTAAGGACTGGGTAGAAA 180

; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-17724

```
Query Match             6.7%;   Score 46.4; DB 10; Length 461;
Best Local Similarity  46.0%;          Pred. No. 0.011;
Matches  143;  Conservative  0;  Mismatches 168;  Indels  0;  Gaps  0;
```

```
Qy   385  AGAATACATGGATGAAAGAAAGGTGAGCACAGCCATTCCTTACAACAGAAATAACAGA   444
          |||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   131  ATAATAAAAAATTTAAAAAAAATAAAAAAAACAACAAGTTGGGACAACATTCCTCTTTATTTT  190

Qy   445  TTTCCCAGCCTCTAACCAAAGAAAAAAAAAAAAAAAAAAATATAAAAAATATTATTTTTTTTTTTTTNNNAAAA  504
          ||||                                                              |||
Db   191  TTTTNAAAAAAAAAAAAAAAATTATTTTTTTTTNNNNAAAAA                         250

Qy   505  GAAAGAGAGGGGATAAATACTGGAGTAGGATGTGAAAAAGTCAAGAGAAAAAAAGA           564
          ||||                                                    ||||||||
Db   251  AAAAAAAAAAAAAAAAAAAATTTAAAAAAAAAAAAAAAAAAAA                        310

Qy   565  ACAGCCCAAGTGTAACAGATACTTCTCCATGGGATGTAAAAAGGAAGTTATTTCAATAA        624
          |||||                                                     |||||
Db   311  AAAAAAAAAAAAAAAAAAAATTTTTAAAAAAAATTTTTTTTTTTTTA                    370

Qy   625  AATGACCCCCTGGAAGGAGTTCAAAACAGGTTCAAACAGGTTGCCACATATGCTTTATGTGGAGTTTTGAAA  684
          |||||                                                                ||||
Db   371  TTTTTTTAAATAAAAAAAAAAAAATTNNNNTTTTTTTTTTTTTTTTTTTTTTTAA             430

Qy   685  AATATATAAAA                                                        695
          || ||||||
Db   431  AAAAAAAAAA                                                         441
```

RESULT 2
US-09-814-353-17724
; Sequence 17724, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17724
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 195, 243, 244, 245, 397, 398, 399

```
Qy   181  AACAAGAGAGTATGAACAAGAGTGAGGAGATTATTAGCAGTGAGTGACCTTTGAGAGTACATCT  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  AACAAGAGAGTATGAACAAGAGTGAGGAGATTATTAGCAGTGAGTGACCTTTGAGAGTACATCT  240

Qy   241  CTAGAGTGGTATGTGAGAGCCAGATTTCAGCATCAGAACCATCAAGCATTTGGGGGT          300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  CTAGAGTGGTATGTGAGAGCCAGATTTCAGCATCAGAACCATCAAGCATTTGGGGGT          300

Qy   301  GGAAGGAAAAGGAGCCATGAATCAAAAGGTGGGGAAAAGGCCTTTGGGAGGTGGCAGTG        360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  GGAAGGAAAAGGAGCCATGAATCAAAAGGTGGGGAAAAGGCCTTTGGGAGGTGGCAGTG        360

Qy   361  TGGGTAGGGAGGTAGGGCTCCGGTTAGAATACATGGATGAAAGAAAGGTGAGCACAGCCAT     420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  TGGGTAGGGAGGTAGGGCTCCGGTTAGAATACATGGATGAAAGAAAGGTGAGCACAGCCAT     420

Qy   421  TTCCTTACAACAGAATAACAGATTTCCCAGCCTCTAACCAAAGAAACAACAAGTTTGG        480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  TTCCTTACAACAGAATAACAGATTTCCCAGCCTCTAACCAAAGAAACAACAAGTTTGG        480

Qy   481  GAACATTCCTCTTCTCTGAAATATGAAAGAGAGGGATAAATACTGGAGTAGGATTGTGA       540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  GAACATTCCTCTTCTCTGAAATATGAAAGAGAGGGATAAATACTGGAGTAGGATTGTGA       540

Qy   541  AAAAAGTCAAGAGAAAAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTCCATGGGATG      600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  AAAAAGTCAAGAGAAAAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTCCATGGGATG      600

Qy   601  GTAAAAAGGAAGTTATTTCAATAAAAATGACCCCCTGGAAGGAGTTCAAAACAGTTGCCA      660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  GTAAAAAGGAAGTTATTTCAATAAAAATGACCCCCTGGAAGGAGTTCAAAACAGTTGCCA      660

Qy   661  TATGCTTTATGTGGAGTTTTGAAAAATATATAAAAT                             696
          |||||||||||||||||||||||||||||||||||
Db   661  TATGCTTTATGTGGAGTTTTGAAAAATATATAAAAT                             696
```

RESULT 3
US-10-311-455-268/c
; Sequence 268, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 268
; LENGTH: 9415
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-268

```
Query Match             6.4%;   Score 44.8; DB 15; Length 9415;
Best Local Similarity  49.2%;          Pred. No. 0.16;
Matches  118;  Conservative  0;  Mismatches 122;  Indels  0;  Gaps  0;
```

```
Qy   387  AATACATGGATGAAAGAAAGGTGAGCACAGCCATTCCTTACAACAGAAATAACAGATT        446
          ||||                                                     |||
Db  7617  AATAACCGTAAAAAAAAATATAAAATCGTAAAAAAAAATTCCCACTAAATAAAAAAAAAA     7558

Qy   447  TCCCAGCCTCTAACCAAAGAAACAACAAGTTGGGACAATTCCTCTTTCTCCGATAAATGA     506
          ||||                                                      |||
Db  7557  ACTTCGCTCCAAACTAAAAAAAAAACACGCGTTAAAAAAACTACCGATAAAAAAAAAC      7498

Qy   507  AAGAGAGGGGATAAATACTGGAGTAGGATTGTGAAAAAGTCAAGAGAAAAAAAAGAGAAC     566
```

```
Db     7497 AAAAAACCGAAAAAAATAAAACCAAAAAATAAAAAAACACCGAATTAAAAAAAAAAAA 7438
                ||| |  ||| |||  |||||||| |  |  |||||||||||||||| |||||||||||
Qy      567 AGCCCAAGTGTAACAGATACTTCTCCATGGGATGGTAAAAAGGAGTTATTTCAATAAAA 626
                ||||  |||||  | ||||||  |||||| |   ||| ||||||||||
Db     7437 CGACCAAAACTAACAACTTCCTAATACTAAAAATTAAATAAAAAAAAAAAACTAACTAAAA 7378
```

; OTHER INFORMATION: EST HUMAN HIT: AW589784.1, EVALUE 1.10e+00
; OTHER INFORMATION: NT HIT: L11910.1, EVALUE 4.00e-02
US-09-864-761-26934

Query Match   6.0%;   Score 41.8; DB 9; Length 300;
Best Local Similarity 52.5%;   Pred. No. 0.17; Indels 2; Gaps 1;
Matches 115; Conservative 0; Mismatches 102; Indels 2; Gaps 1;

```
Qy        3 AAATATCTGATAGGCAGTTAGAAATTGAGTTTGGAACACAGAGGAGAGGCTTTGATGGC 62
               |||  | ||||  ||||||| |||||||||||||||
Db      228 AAATGTGAAGTTTGCAGTTGAATATTGTGCCTGGAACTTAGGAAAAACAATGTGTGGGC 169
Qy       63 GA--TACAGACTTGGAAGACATCAGTGCTGAGCAGTAAATGAGTGATTCAGGAAAGAGT 120
               ||   | |||||  || ||||||  | ||||
Db      168 CATCTGCATATCGGGTGATGAATGAAGCCACAGAGCAGAGACAGAGCAGCAAGAGAGT 109
Qy      121 ATAAACTGGAAGAGGACAGAGGACAGGCTCAAGGAACTCAAGGAAGGACATATTTAAGGACTGGGTAGAAA 180
               ||||   |  ||||||||
Db      108 AAATAAAGTGAAGAAGGAGGGGAAGAGGAAGGACAGAAGAGGAGGAAGCCAATGTTTAAGGG 49
Qy      181 AACAAGAGTATGAACAAGAGTGAGGAGATTATTAGCA 219
               || ||| || ||||||| |||
Db       48 AATGATAGTGTATTCAAAATTGGCAAAGAGACTGAGAA 10
```

RESULT 5
US-09-814-353-18511/c
; Sequence 18511, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18511
; LENGTH: 508
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-18511

Query Match   5.9%;   Score 41; DB 10; Length 508;
Best Local Similarity 52.7%;   Pred. No. 0.39; Indels 0; Gaps 0;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy      399 AAAGAAAGGTGAGCAGCACAGCCATTCCTTTACAACAGAATAACAGATTTCCCAGCCTCTA 458
               |||||||| |||  |   ||||||||||||  |||||     | |  ||| ||||  ||||
Db      280 AAAAAAAAGGTTTTTACAACAATTTTTTTTTTCCCAAAAAAAGTTTTTTCCCCCCCAAAA 221
Qy      459 ACCAAAGAAACAACAAGTTTGGGAACATTCCTCCTTCTGAAATATGAAAGAGAGGGGGAT 518
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      220 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAAAAAAAAAAAAAAAAAAAAAAAAA 161
Qy      519 AAATACTGGAGTAGATTGTGAAAAAGTCAAGAGAAAAAGAGAACA 567
               ||| ||||   ||| |||||   |||||||||
```

RESULT 4
US-09-864-761-26934/c
; Sequence 26934, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24463.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26934
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012079.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.99
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.97

Db   160 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAA 112

RESULT 6
US-10-311-455-1669/c
; Sequence 1669, Application US/10311455
; Publication No. US2003014360GA1
; GENERAL INFORMATION:
; APPLICANT: OLEK,Alexander
; APPLICANT: PIEPENBROCK,Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1669
; LENGTH: 6668
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 4733
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1669

Query Match   5.8%; Score 40.6; DB 15; Length 6668;
Best Local Similarity 50.3%; Pred. No. 2;
Matches 100; Conservative   0; Mismatches 99; Indels   0; Gaps   0;

Qy   428 ACAACAGAAATAACAGATTTCCCAGCCTCTAACCAAAGAAACAACAAGTTTGGGAACATT 487
              |||| ||||| || |||||||||| |||| |||| ||||| ||| |||
Db   5488 ACGAAAAAAAAAAAACAAAACTTTCTAAAAAAAAAAAAAAAAACTTCCAAAACCAAT 5429

Qy   488 CCTCTCTTCTGAAATATGAAAGAGAGGGGATAAATACTGGAGTAGGATTGTGAAAAAAGT 547
              ||| |||| |||||||||||||||||| |||||||| ||||||||||||
Db   5428 CTTACCAAAT AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5369

Qy   548 CAAGAGAAAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTCCCATGGGATGGTAAAAA 607
              ||| ||||||||| ||| ||| |||| ||||||
Db   5368 AAAAAAAAAAAAAAAAAAAAAAAAAAACGAACGAAAAAAAAAAAAAAAA 5309

Qy   608 GGAAGTTATTTCAATAAAA 626
              |||| ||| ||||
Db   5308 CAAAACTTTCTAAAAAAAA 5290

RESULT 7
US-10-184-644-346
; Sequence 346, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 346
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-346

Query Match   5.8%; Score 40.2; DB 15; Length 671;
Best Local Similarity 8.6%; Pred. No. 0.76;
Matches 46; Conservative 172; Mismatches 317; Indels   0; Gaps   0;

Qy   34 TTGGAACACAGGAGAGAGGGCTTTGATGGCGATACAGACTTGGAAGACATCAGTGCTGAGC 93
            : :   :: : :: | : ::   :: : :: : :::
Db   70 KPNKRKGFNEGLWEIQNNPHASYSAPPPVSSSDSEAPEANPADGSDADEDDEDRGVMAVT 129

Qy   94 AGTAAATGAGATGATTCAGGAAAGAGTATAAACTGGGAAGAGGACAGAGGACAGGCTCAA 153
            : :: :::: ::: :::::: :   :   ::::::::: :::::: :::::::
Db   130 AVTATAASDRMESDSDSDKSSDNSGLKRKTPALKMSVSKRARKASSDLDQASVSPSEEEN 189

Qy   154 GGAACATATTTAAGGACTGGGTAGAAAAACAAGAGTATGAACAAGAGTGAGGAGATTA 213
            : : ::   ::: : : :: ::   :| :: |:: :: ::: :::::
Db   190 SESSESEKTSDQDFTPEKKAAVRAPRRGPLGGRKKKKAPSASDSDSKADSDGAKPEPVA 249

Qy   214 TTAGCAGTGACCTTTCGAGTACATCTCAGAGTGTATGTGTAGAGCCAGATTTTCAGC 273
            : ::::: ::::: :: : :::: : ::::: :: :: :
Db   250 MARSASSSSSSSSDSDVSVKKPPRGRKPAEKPLPKPRGRKPKPERPPSSSSDSDSDE 309

Qy   274 ATCAGAACCATCAAGCATTTTGGGGTGGAAGGAAAAGGAGCCATGAATCAAACAAAGGTGGG 333
            :: : ::::: :: : : :::::::::::::: ::::: |:
Db   310 VDRISEWKRDEARRRELEARRRREQEEELRRLREQEKEEKERRRERADRGEAERGSGS 369

Qy   334 GAAAAGGCCTTTGGGAGGTGGCAGTGTGGGTAGGGAGTAGGGCTCCGGTTAGAATACAT 393
            : : | ::::::|::| |:|: ||:: ::
Db   370 SGDELREDDEPVKKRGRKGRGRGPSSSDSEPEAELEREAKKSAKKPQSSSTEPARKPGQ 429

Qy   394 GGATGAAAGAGAAGGTGAGCACAGCCATTCCTTTACAACAGAAATAACAGATTTCCCAGC 453
            |||::::| |:::::: | :::::::| :::: |||| ||||| |:::||
Db   430 KEKRVRPEEKQQAKPVKVERTRKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVD 489

Qy   454 CTCTAACCAAAGAAACAACAAGTTTGGGAACATTCCTCCTTCTGAAATATGAAAGAGAG 513
            : : |||| ::|:|| : || |||| ::::|
Db   490 SPDVKRCLNALEELGTLQVTSQILQKNTDVVATLKKIRRYKANKDVMEKAAEVYTRLKSR 549

Qy   514 GGGATAAATACTGGAGTAGGATTGTGAAAAAAGTCAAGAGAAAAAAAAAGAACAG 568
            |||: :: ||||:::   :::   :: :| :|
Db   550 VLGPKIEAVQVKNKAGMEKEKAEEKLAGEELAGEEAPQEKAEDKPSTDLSAPVNG 604

RESULT 8
US-10-184-634-346
; Sequence 346, Application US/10184634
; Publication No. US2003006868A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634

```
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 346
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-346

Query Match             5.8%;  Score 40.2; DB 15; Length 671;
Best Local Similarity   8.6%;  Pred. No. 0.76;
Matches   46;  Conservative  172;  Mismatches  317;  Indels   0;  Gaps   0;

Qy    34 TTGGAACACAGGAGAGAGGCTTTGATGCGATACAGACTTGGAAGACATCAGTGCTGCAGC 93
                                      :: ::   ::::::     :  :::  ::
Db    70 KPNRKRGFNEGLWEIQNNPHASYSAPPVSSSDSEAPEANPADGSDADEDDEDRGVMAVT 129

Qy    94 AGTAAATGAGATGATTCAGGAGGAAGAGAGTATAAACTGGGAAGGACACAGGCTCAA 153
         |:||| | ::::: :::: |:  ::::  :  |::::: | :::::: |::::::
Db   130 AVTATAASDRMESDSDSDKSSDNSGLRKTPALKMSVSKRARKASSDLDQASVSPSEEEN 189

Qy   154 GGAACATATTTAAGGACTGGGTAGAAAAACAAGAGAGTATGAACAAGAGTGAGGAGATTA 213
         ::  ::   ::  :: :::::  ::: ::  :: :::::
Db   190 SESSSESEKTSDQDFTPEKKAAVRAPRGRGPLGRKKKKAPSASDSDSKADSDGAKPEPVA 249

Qy   214 TTAGCAGTGACCTTTGAGAGTACATCTCTAGAGTGGTATGTGTAGAGCCAGATTTTCAGC 273
         :::::       :::::::  :::: ::  : :::: :::::: ::::   :   :::
Db   250 MARSASSSSSSSSSDSDVSVKKPPRGRKPAEKPLPKPGRGRKKPERPPSSSSSDSDSDE 309

Qy   274 ATCAGAACCATCAGCATTTTGGGGGTGGAAGGAAGGAGGCCATGAATCAAAAGGTGGG 333
         ::  ::   :::: ::  ::::  :::: :::  :: :::
Db   310 VDRISEWKRDEARRRELEARRRREQEEELRLRLREQEKEEKERRRERADRGEAERGSGGS 369

Qy   334 GAAAAGGCCTTTGGAGGTGGCAGTGCTGGGTAGGGAGTGGGTAGGGCTCCGGTTAGAATACAT 393
         :::: ::||::|:  :::::::
Db   370 SGDELREDDEPVKKRGRKRGRGRGPPSSSDSEPEAELEREAKKSAKKPQSSSTEPARKPGQ 429

Qy   394 GGATGAAAGAAAGGTGAGCACAGCCATTTCCTTACAACAGAAATAACAGATTTCCCAGC 453
         ::   ::  ::   ::  :::  :: ::::  ::
Db   430 KEKRVRPEEKQQAKPVKVERTRKRSEGFSMDRKVEKKKEPSVEEKLQKLHSEIKFALKVD 489

Qy   454 CTCTAACCAAAGAACAACAAGTTGGGAACATTCCTCTTCCTTCTGAAATATGAAAGAGAG 513
         ::  ::   : ::: :::::::
Db   490 SPDVKRCLNALEELGTLQVTSQILQKNTDVVATLKKIRRYKANKDVMEKAAEVYTRLKSR 549

Qy   514 GGGATAAATACTGGAGTAGGATTGTGAAAAAGTCAAGAGAAAAAAAAGAACAG 568
         :: : ::   : :::::: :::: ::: :   ::
Db   550 VLGPKIEAVQKVNKAGMEKEKAEEKLAGEELAGEEAAPQEKAEDKPSTDLSAPVNG 604
```

RESULT 9
US-10-282-122A-16543
; Sequence 16543, Application US/10282122A
; Publication No. US2004002912A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

```
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 16543
; LENGTH: 1269
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-282-122A-16543

Query Match             5.8%;  Score 40.2; DB 13; Length 1269;
Best Local Similarity  47.6%;  Pred. No. 1.1;
Matches  151;  Conservative    0;  Mismatches  163;  Indels   3;  Gaps   1;

Qy    19 GTTAGAAATTTGAGTTTGGAACACAGGAGAGAGGCTTGATGGCGATACAGACTTCGAAG 78
                     ::::::: ::::::::::::  ::::::
Db   400 GCTTTAAATATAGGCTATAAACATTCAAGTGGAGACTTTATAGCCGTATATGATCGACGAT 459

Qy    79 ACATCAGTGCTGAGCAGTAAAATGAGATGATTCAGGAGAGAGATTAAAACTGGGAGGAGGAC 138
                 ||||||
Db   460 AATACTTCCAGATAAAAAATGCATTAAAATGTCATTATGTGAAAATCTATATATTGAAAAAC 519

Qy   139 AGAGGACAGGCTCAAGGAGCTCAAGGAACATATTTAAGGACTGGGTAGAAAAACAAGAGAGTATGAACA 198
              ::::::::::
Db   520 TTAGGAGCAGTTATAGGAA---AGTTTAGAACTGAAAATAAAGACACAGAAATATGTTAACA 576

Qy   199 AGAGTGAGGAGGATTATTAGCAGTGACCTTTGAGAGTACATCTCTAGAGTGTATGTGTAG 258
                  ::::::::
Db   577 AGATTTATAAATATAGAGACACCTTAAGTTTTCAATGGATGTGCCAGCAGGAGATGGAAG 636

Qy   259 AGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGTGGAAGGAGAAAGAGAGGCCAT 318
                   :::::::||||
Db   637 TTATTAAATTTATGTACTATACCTGCACTAATTTGTAGTAAGAAGAAGAATATATACAA 696

Qy   319 GAATCAAAAGTGGGGA 335
         ||||  | ::||||
Db   697 GAGTTGAATGGATGGGA 713
```

RESULT 10
US-10-027-632-153507/c
; Sequence 153507, Application US/10027632
; Publication No. US2002019371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29

Query Match   5.7%;   Score 40; DB 16; Length 767;
Best Local Similarity 49.3%; Pred. No. 0.93;
Matches 133; Conservative 0; Mismatches 135; Indels 2; Gaps 1;

```
Qy  381  GGTTAGAATACATGGATGAAAGAAAGGTGAGCACAGCCATTTCCTTTACAACAGAAATAA  440
         ||| |||||  |||| |||||||  ||| |||||||| ||||| ||
Db  651  GGCTATAATATGTTGATCAATCAGTTGTAAATGCAGTCATTTGGA--GAAACAGC  594

Qy  441  CAGATTTCCCAGCCTCTAACCAAAGAAACAACAAGTTGGGAACATTCCTCTTCTTCTGAA  500
         ||||||||||||| ||||||||||||||||||||||||||||||||
Db  593  CCTCTTTCTGGTGCTGTTTACAAAGAGTAGTGACCCATTCTTAAACCTGGG  534

Qy  501  ATATGAAAGAGAGGGGATAAATACTGGAGTAGGATTGTGAAAAAGTCAAGAGAAAAAA  560
         ||| | ||||||||||||||||
Db  533  ACATAAAAAGGACCGTATATATGTGAATAATGGTAGGAGAAAAATCACCTATAGAAAAG  474

Qy  561  AAGAACAGCCCAAGTGTAACAGTACTTCTCTGTAAAGGAGTTCAAAA  620
         ||||||||||
Db  473  AATGGAAGTCAAGTTCCTCGTGTAATGTATGATGCTTGATTTCAAGTGAATTTAGCCA  414

Qy  621  ATAAAATGACCCCTTGGAAGGAGTTCAAAA  650
         ||||| ||||||||
Db  413  ATAAAGTCCATGCCCTTAAGAAGTTTATAA  384
```

RESULT 12
US-10-311-455-1616/c
; Sequence 1616, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1616
; LENGTH: 6319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1616

Query Match   5.7%;   Score 39.6; DB 15; Length 6319;
Best Local Similarity 48.6%; Pred. No. 3.8;
Matches 108; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

```
Qy  427  TACAACAGAAATAACAGATTTCCCAGCCTCTAACCAAAGAAACAACAAGTTTGGGAACAT  486
         ||||||| |||||||  ||  |||||||||     || |
Db  705  TAAAAAACATAAAAAAATAAAAAAAAAAACTAAAAAAAAAAACTAAAATTAAAAT  646

Qy  487  TCCTCTCTTCTGAAATATGAAAGAGAGGGGATAAATACTGGAGTAGGATTGTGAAAAAAG  546
         | |  ||  | |||||||||    ||| ||||||
Db  645  ACAAATCTTCTTACGAAATATAATATAAAAAAAAATACCACAAATAAAAATAAAAATT  586

Qy  547  TCAAGAGAAAAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTCCATGGGATGGTAAAA  606
         ||| | ||||||||||||||||
Db  585  AAAAACTAAAAAAACTAAAAACAAAACAAAAAAAATTCAAAAAACTAAAAAAACA  526

Qy  607  AGGAAGTTATTTCAATAAAAATGACCCCCTTGGAAGGAGTTCAA  648
         |||||| ||||||||
Db  525  ACTAACTAATTCCTAAAAACTCTTCCCTTATCATCTCATCAA  484
```

; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR APPLICATION FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR APPLICATION FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR APPLICATION FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR APPLICATION FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 153507
; LENGTH: 767
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-153507

Query Match   5.7%;   Score 40; DB 13; Length 767;
Best Local Similarity 49.3%; Pred. No. 0.93;
Matches 133; Conservative 0; Mismatches 135; Indels 2; Gaps 1;

```
Qy  381  GGTTAGAATACATGGATGAAAGAAAGGTGAGCACAGCCATTTCCTTTACAACAGAAATAA  440
         ||| |||||  |||| |||||||  ||| |||||||| ||||| ||
Db  651  GGCTATAATATGTTGATCAATCAGTTGTAAATGCAGTCATTTGGA--GAAACAGC  594

Qy  441  CAGATTTCCCAGCCTCTAACCAAAGAAACAACAAGTTGGGAACATTCCTCTTCTTCTGAA  500
         ||||||||||||| ||||||||||||||||||||||||||||||||
Db  593  CCTCTTTCTGGTGCTGTTTACAAAGAGTAGTGACCCATTCTTAAACCTGGG  534

Qy  501  ATATGAAAGAGAGGGGATAAATACTGGAGTAGGATTGTGAAAAAGTCAAGAGAAAAAA  560
         ||| | ||||||||||||||||
Db  533  ACATAAAAAGGACCGTATATATGTGAATAATGGTAGGAGAAAAATCACCTATAGAAAAG  474

Qy  561  AAGAACAGCCCAAGTGTAACAGTACTTCTCTGTAAAGGAGTTCAAAA  620
         ||||||||||
Db  473  AATGGAAGTCAAGTTCCTCGTGTAATGTATGATGCTTGATTTCAAGTGAATTTAGCCA  414

Qy  621  ATAAAATGACCCCTTGGAAGGAGTTCAAAA  650
         ||||| ||||||||
Db  413  ATAAAGTCCATGCCCTTAAGAAGTTTATAA  384
```

RESULT 11
US-10-027-632-153507/c
; Sequence 153507, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 153507
; LENGTH: 767
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-153507

RESULT 13
US-09-854-867-362/c
; Sequence 362, Application US/09854867
; Publication No. US2003224356A1
; GENERAL INFORMATION:
;   APPLICANT: JOAN, KNOLL H
;   APPLICANT: ROGAN, PETER K
;   TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
;   FILE REFERENCE: 30307
;   CURRENT APPLICATION NUMBER: US/09/854,867
;   CURRENT FILING DATE: 2003-05-08
;   NUMBER OF SEQ ID NOS: 613
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 362
;   LENGTH: 2750
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: (1)..(2750)
;   OTHER INFORMATION: line2
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (17)..(17)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (89)..(89)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (116)..(116)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (433)..(433)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (500)..(500)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (548)..(548)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (578)..(578)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (604)..(604)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (709)..(709)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (772)..(772)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (797)..(797)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1117)..(1118)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature

; LOCATION: (2099)..(2099)
; OTHER INFORMATION: n is a, c, g or t
US-09-854-867-362

Query Match   5.7%;   Score 39.4; DB 13; Length 2750;
Best Local Similarity 67.0%;   Pred. No. 2.7;
Matches   69; Conservative   1; Mismatches 32; Indels   1; Gaps   1;

Qy   3 AAATATCTGATAGGCAGTTAGAAATTTGAGTTTGGAACACCAGGAGAGAGG-CTTTGATGG 61
               ||||||||| ||| |||||||||||| || ||: |||||||:| ||||||| ||||||||
Db   1743 AAATGTCGAGTAGGCAGTTCGGATATAGGAGTCTCGGARTTCAGGGGAGAGTCTCGGGCTGG 1684

Qy   62 CGATACAGACTTGGAAGACATCAGTGCTGCTGAGCAGTAAATGAGA 104
               || ||| ||||||||| || || | ||:|||||||||||||||
Db   1683 AGATATAAATTGGGAGTCATCAGCGTATAGCGTATAGATGATGTATTTAAA 1641

RESULT 14
US-10-311-455-128/c
; Sequence 128, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;   TITLE OF INVENTION: cytosine methylation
;   FILE REFERENCE: 5013.1014
;   CURRENT APPLICATION NUMBER: US/10/311,455
;   CURRENT FILING DATE: 2002-12-16
;   PRIOR APPLICATION NUMBER: PCT/EP01/07537
;   PRIOR FILING DATE: 2001-07-02
;   PRIOR APPLICATION NUMBER: DE 10032529.7
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: DE 10043826.1
;   PRIOR FILING DATE: 2000-09-01
;   NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 128
;   LENGTH: 6681
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-128

Query Match   5.7%;   Score 39.4; DB 15; Length 6681;
Best Local Similarity 46.3%;   Pred. No. 4.4;
Matches 130; Conservative   0; Mismatches 151; Indels   0; Gaps   0;

Qy   413 ACAGCCATTCCTTTACAACAGAAATAACAGATTTCCAGCCTCTAACCAAAGAAACAAC 472
               ||||||||||||| |||||||||||||||||| ||||||||||||||||||
Db   5867 ATATCTCTTTTCTTTAAAAACACAAAAACAAATAATTAACTAAAARTAAAARTAAAATAC 5808

Qy   473 AAGTTGGGAACATTCCTCCTTCTGAAATATGAAAGAGAGGGGATAAATACTGGAGTAG 532
               ||||||||||| ||||| | ||||||| |||||| |||||||||||||||||||
Db   5807 AATAAAAAAATTAAAAAAATTAAAAAATTTAAAAAAATACATAAAACATAAAACTAA 5748

Qy   533 GATTGTGAAAAAGTCAAGAGAAAAAAAAAGAACAGCCCAGTGTAACAGATACTTCTCC 592
               |||||| |||| ||| ||||| |||||||| || ||||||||| |||||||||||
Db   5747 AAAACTAAACCTATTACTAAAAAATAACAAAAAAAATCTTAAATAATTACTAATCCCAATCT 5688

Qy   593 ATGGGATGGTAAAAGGAAGTTATTCCATAAAATGACCCCTTGGAAGGAGTTCAAAACA 652
               ||||||| ||||||||||||||||||| ||| | ||||| || | ||||||
Db   5687 ACAAAATTACATCTAAATTATATATTATATATTTATTATATAATTAAAATATATTA 5628

Qy   653 GGTTGCCATATGCTTTATGTGGAGTTTGAAAATATATAA 693
               ||||||||||||||||||||||||||||||||||
Db   5627 ACATCTTAACATATTTATATATTTATTATATAATTATTA 5587

RESULT 15
US-10-240-452-4/c
; Sequence 4, Application US/10240452

; Publication No. US20030162194A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Apoptosis
; FILE REFERENCE: 5013.1006
; CURRENT APPLICATION NUMBER: US/10/240,452
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03969
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 4
;   LENGTH: 6681
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-452-4

Query Match   5.7%;   Score 39.4; DB 15; Length 6681;
Best Local Similarity 46.3%;   Pred. No. 4.4;
Matches 130; Conservative   0; Mismatches 151; Indels   0; Gaps   0;

Qy   413 ACAGCCATTTCCTTTACAACAGAAATAACAGATTTCCCAGCCTCTAACCAAAGAAACAAC 472
          || ||||| || ||||||| || ||||| | || |||| ||| ||||||| || ||
Db   5867 ATATCTCTTTTCTTTAAAAACAAAAACAAAATAATTAACTAAAAATAAAATAAAATAC 5808

Qy   473 AAGTTTGGGAACATTCCTCTCTTCTGAAATATGAAAGAGAGGGGATAAATACTGGAGTAG 532
          || | | ||| | || ||| |||| |||| |||||| ||| | ||||| ||| |
Db   5807 AATAAAAAATTTAAAAAAATATTAAAAAATTTAAAAAAATTAAAAAATACATAAAACTAA 5748

Qy   533 GATTGTGAAAAAAGTCAAGAGGAAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTCC 592
          ||| || || ||| |||| | |||| || ||||| || || ||| | ||| || ||
Db   5747 AAAACTAAACCTATTACTAAAAAAATAACAAAAAATCTTAAATAATAACTAATCCCAATCT 5688

Qy   593 ATGGGATGGTAAAAGGAAGTTATTCAATAAAATGACCCCTTGGAAGGAGTTCAAAACA 652
          ||| | || | ||||| |||| || | ||| | ||||| ||| ||| |||||
Db   5687 ACAAAATTACATCTAAATATTAAAATAATTAAAACAAAAATTAAAAATATCTAAAAATTA 5628

Qy   653 GGTTGCCATATGCTTTATGTGGAGTTTGAAAAATATATAA 693
          |||| ||| ||| | | | ||||| | |||||| ||||
Db   5627 ACATCTTAACATATTTATATATATATTATTATATAATTATTA 5587

Search completed: April 26, 2004, 03:19:16
Job time : 207.468 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:15:18 ; Search time 34.5277 Seconds
                                           (without alignments)
                                           11186.557 Million cell updates/sec

Title: US-10-051-835-14
Perfect score: 696
Sequence: 1 taaaatatctgataggcagt..........ttttgaaaatatataaaat 696

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
                1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
0	1	89	12.8	7218	1	US-08-232-463-14	Sequence 14, Appl
0	2	40.8	5.9	6305	3	US-08-687-580B-6	Sequence 6, Appl
0	3	37.8	5.4	604	4	US-09-334-818A-8	Sequence 8, Appli
	4	37	5.3	289	3	US-09-007-005-17	Sequence 17, Appl
	5	37	5.3	289	3	US-09-244-796-17	Sequence 17, Appl
	6	36.4	5.2	2089	1	US-08-552-142A-1	Sequence 1, Appl
0	7	36.4	5.2	2089	1	US-08-910-973-1	Sequence 1, Appl
0	8	36.4	5.2	2089	4	US-09-499-227-1	Sequence 1, Appli
0	9	36.4	5.2	2089	5	PCT-US95-05741-1	Sequence 1, Appli
0	10	36.2	5.2	832	3	US-09-621-976-2813	Sequence 2813, Ap
0	11	36.2	5.2	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	12	36	5.2	1821	4	US-08-477-831C-1	Sequence 1, Appli
	13	36	5.2	1885	4	US-08-477-831C-9	Sequence 9, Appli
	14	36	5.2	1896	4	US-08-477-831C-10	Sequence 10, Appl
	15	36	5.2	1961	4	US-08-477-831C-8	Sequence 8, Appli
	16	36	5.2	2968	4	US-08-477-831C-13	Sequence 13, Appl
	17	36	5.2	3044	4	US-08-477-831C-12	Sequence 12, Appl
	18	36	5.2	392000	4	US-10-027-983-11	Sequence 11, Appl
0	19	35.8	5.1	3084	4	US-09-620-312D-616	Sequence 616, App
0	20	35.4	5.1	505	4	US-09-621-976-15639	Sequence 15639, A
0	21	35.4	5.1	6124	3	US-08-213-419B-3	Sequence 3, Appli
	22	35	5.0	248	3	US-09-007-005-32	Sequence 32, Appl
	23	35	5.0	248	3	US-09-244-796-32	Sequence 32, Appl
	24	35	5.0	277	3	US-09-007-005-3	Sequence 3, Appli
	25	35	5.0	277	3	US-09-244-796-3	Sequence 3, Appli
	26	35	5.0	1830121	4	US-09-557-884-1	Sequence 1, Appli
	27	35	5.0	1830121	4	US-09-643-990A-1	Sequence 1, Appli

0	28	34.8	5.0	1166	4	US-09-072-596-323	Sequence 323, App
0	29	34.8	5.0	1166	4	US-09-072-967-328	Sequence 328, App
0	30	34.8	5.0	2502	3	US-09-234-332-1	Sequence 1, Appli
	31	34	4.9	2653	3	US-08-589-711-1	Sequence 1, Appli
	32	34	4.9	2653	3	US-09-221-938-1	Sequence 1, Appli
	33	34	4.9	2653	3	US-08-945-476-7	Sequence 7, Appli
	34	34	4.9	640681	4	US-09-790-988-1	Sequence 1, Appli
0	35	34	4.9	640681	4	US-09-790-988-1	Sequence 45, Appl
0	36	33.6	4.8	395	4	US-10-204-708-60	Sequence 60, Appl
0	37	33.6	4.8	6156	4	US-08-559-896B-3	Sequence 3, Appli
	38	33.4	4.8	417	4	US-09-351-794A-3	Sequence 3, Appli
	39	33.4	4.8	417	4	US-08-559-896B-1	Sequence 1, Appli
	40	33.4	4.8	1956	4	US-09-351-794A-1	Sequence 1, Appli
	41	33.4	4.8	1956	4	US-09-976-594-592	Sequence 592, App
0	42	33.4	4.8	4557	4	US-09-621-976-18033	Sequence 1803, A
0	43	33.2	4.8	474	1	US-08-340-539A-11	Sequence 11, Appl
0	44	33.2	4.8	1696	1	US-08-461-592B-11	Sequence 11, Appl
0	45	33.2	4.8	1696	2		

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No.5670367
;-- GENERAL INFORMATION:
;-- APPLICANT: DORNER, F.
;-- APPLICANT: SCHEIFLINGER, F.
;-- APPLICANT: FALKNER, F. G.
;-- TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;-- NUMBER OF SEQUENCES: 52
;-- CORRESPONDENCE ADDRESS:
;-- ADDRESSE: Foley & Lardner
;-- STREET: 1800 Diagonal Road, Suite 500
;-- CITY: Alexandria
;-- STATE: VA
;-- COUNTRY: USA
;-- ZIP: 22313-0299
;-- COMPUTER READABLE FORM:
;-- MEDIUM TYPE: Floppy disk
;-- COMPUTER: IBM PC compatible
;-- OPERATING SYSTEM: PC-DOS/MS-DOS
;-- SOFTWARE: PatentIn Release #1.0, Version #1.25
;-- CURRENT APPLICATION DATA:
;-- APPLICATION NUMBER: US/08/232,463
;-- FILING DATE:
;-- CLASSIFICATION: 435
;-- PRIOR APPLICATION DATA:
;-- APPLICATION NUMBER: US/07/935,313
;-- FILING DATE:
;-- APPLICATION NUMBER: EP 91 114 300.6
;-- FILING DATE: 26-AUG-1991
;-- ATTORNEY/AGENT INFORMATION:
;-- NAME: BENT, Stephen A.
;-- REGISTRATION NUMBER: 29,768
;-- REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;-- TELECOMMUNICATION INFORMATION:
;-- TELEPHONE: (703)836-9300
;-- TELEFAX: (703)683-4109
;-- TELEX: 899149
;-- INFORMATION FOR SEQ ID NO: 14:
;-- SEQUENCE CHARACTERISTICS:
;--   LENGTH: 7218 base pairs
;--   TYPE: nucleic acid
;--   STRANDEDNESS: single
;--   TOPOLOGY: linear
;-- IMMEDIATE SOURCE:
;--   CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match   12.8%; Score 89; DB 1; Length 7218;

Best Local Similarity 4.3%; Pred. No. 5.5e-16;
Matches   17; Conservative 249; Mismatches 129; Indels   0; Gaps   0;

```
Qy      13 TAGGCAGTTAGAAATTGAGTTGGAACACAGGAGAGAGGCTTTGATGGCGATACAGACT   72
           ||||  | |||| ||||||| || ||||| ||       :: :: :::
Db    1462 TAGTTAAAGAGGATAGAGAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1403

Qy      73 TGGAAGACATCAGTGCTGAGCAGTAAATGAGATGATTCAGGAAAGAGTATAAACTGGGAA  132
           :: :: : :: ::: :  :::::::::: :  :::::
Db    1402 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1343

Qy     133 GAGGACGAGAGGACAGAGGCTCAAGGAACATATTTAAGGACTGGGTAGAAAAACAAGAGAGTA  192
           ::::: ::::: ::: :::::  :::  ::::::::: ::::::::
Db    1342 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1283

Qy     193 TGAACAAGAGTGAGGAGATTATTAGCAGTGACCTTTGAGAGTACATCTCTAGAGTGGTAT  252
           :: :::::  ::::::: ::  :: : ::::: :::::::: :::: :::
Db    1282 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1223

Qy     253 GTGTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGGTGGAAGGAAAAGG  312
           : : :::::: ::::: ::::: :::::::::::::::::::::::
Db    1222 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1163

Qy     313 AGCCATGAATCAAAAGGTGGGGAAAAGGCCCTTTGGGAGGTGGCAGTGTCGGGTAGGGAGT  372
           ::  :: ::::::: :::::::::  :::::: :: ::::::
Db    1162 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1103

Qy     373 AGGGCTCCGGTTAGAATACACATGGATGAAAGAAAGG  407
           ::::::::::::::: ::::::::::::::::
Db    1102 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1068
```

RESULT 2
US-08-687-580B-6/c
; Sequence 6, Application US/08687580B
; Patent No. 6291647
; GENERAL INFORMATION:
; APPLICANT: MELCHERS, Leo Sjoerd
; APPLICANT: PONSTEIN, Anne Silene
; APPLICANT: KROON-SWART, Saskia
; APPLICANT: VAN DEVENTER-TROOST, Johanna Petronella Els
; APPLICANT: OHL, Stephan Andres
; APPLICANT: BRES-VLOEMANS, Alexandra Aleida
; APPLICANT: LOGEMANN, J rgen
; APPLICANT: SELA-BUURLAGE, Marianne Beatrix
; TITLE OF INVENTION: Antifungal proteins, DNA coding therefor, and
; TITLE OF INVENTION: hosts incorporating same
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: LADAS & PARRY
;   STREET: 26 WEST 61 STREET
;   CITY: NEW YORK
;   STATE: NY
;   ZIP: 10023
;   COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3-1/4" Disk 1.44MB
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: Microsoft Windows for Workgroups 3.11
;   SOFTWARE: WordPerfect 8 for Windows
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/687,580B
;   FILING DATE: 20-NOV-1996
;   CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/EP95/00488
;   FILING DATE: 2-SEP-1994
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: NL94200321.1
;   FILING DATE: 9-FEB-1994
; ATTORNEY/AGENT INFORMATION:
;   NAME: MASS, Clifford J.
;   REGISTRATION NUMBER: 30,086

; REFERENCE/DOCKET NUMBER: U-010885-8
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 708 1890
;   TELEFAX: (212) 246 8959
;   TELEX: 233288
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 6305 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
;   ORGANISM: Nicotiana tabacum
;   STRAIN: Samsun NN
;   TISSUE TYPE: leaf
; IMMEDIATE SOURCE:
;   CLONE: lambda 3
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 774..911
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 1045..1154
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 3105..3375
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 4226..4328
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 4432..4777
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 4891..5253
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 5336..5414
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 5531..5743
; FEATURE:
;   NAME/KEY: intron
;   LOCATION: 912..1044
; FEATURE:
;   NAME/KEY: intron
;   LOCATION: 1155..3104
; FEATURE:
;   NAME/KEY: intron
;   LOCATION: 3376..4225
; FEATURE:
;   NAME/KEY: intron
;   LOCATION: 4329..4431
; FEATURE:
;   NAME/KEY: intron
;   LOCATION: 4778..4890
; FEATURE:
;   NAME/KEY: intron
;   LOCATION: 5254..5335
; FEATURE:
;   NAME/KEY: intron
;   LOCATION: 5415..5530
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: join(774..911, 1045..1154, 3105..3375, 4226..4328,
;   LOCATION: 4432..4777, 4891..5253, 5336..5414, 5531..5743)
; FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: 774..839
; FEATURE:
;   NAME/KEY: polyA_site
;   LOCATION: 6095..6100

US-08-687-580B-6

```
; OTHER INFORMATION:  /label= putative

Query Match      5.9%;   Score 40.8;  DB 3;  Length 6305;
Best Local Similarity   54.7%;  Pred. No. 0.065;
Matches    81;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy      72 TTGGAAGACATCAGTGCTGAGCAGTAAATGAGATGATTCAGGAAGAGTATAAACTGGGA 131
           |||  ||  |||| ||  |||| || |||||  |  || |||| |||||  ||||| |||
Db    3093 TTCCAAGATGTCAATTTAAATCTTTAAATAAGTAATTCAGAAAAATATTGTAATATAGTA 3034

Qy     132 AGAGGACAGAGGACAGGCTCAAGGAACATATTTAAGGACTGGGTAGAAAAACAAGAGAGT 191
           ||||| ||  | ||  ||   ||  |||||||  || ||   ||    ||| ||||||||
Db    3033 GAAGGGCATTGATTTATCTAATATTTAATATTTAATAAGATTGTAGGGTCAGAAAATAAT 2974

Qy     192 ATGAACAAGAGTGAGGAGATTATTAGCA 219
           |||   |||||| ||  ||| |||| |||
Db    2973 AGAAACAAAAGAGCTGAAATGGCCAACA 2946
```

RESULT 3
US-09-334-818A-8/c
; Sequence 8, Application US/09334818A
; Patent No. 6479267
; GENERAL INFORMATION:
; APPLICANT: DAVIS, MARIA
; APPLICANT: FULLER, CARL W.
; APPLICANT: MAMONE, JOSEPH A.
; APPLICANT: HUANG, LIN
; TITLE OF INVENTION: FY7 POLYMERASE
; FILE REFERENCE: PB9817
; CURRENT APPLICATION NUMBER: US/09/334,818A
; CURRENT FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: 60/089,556
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Thermus sp.
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (266)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (293)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (299)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (301)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (304)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (312)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (332)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (335)..(338)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (341)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (343)..(346)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (349)..(350)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (352)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (358)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (363)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (366)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (372)..(375)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (379)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (385)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (391)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (393)..(395)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (400)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (403)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (406)..(408)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (422)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
```

```
; NAME/KEY: modified_base
; LOCATION: (428)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (432)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (434)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (436)..(438)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (440)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (442)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (448)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (455)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (458)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (469)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (479)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (484)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (487)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (491)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (493)..(494)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (528)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (535)..(537)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (539)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (551)..(552)
; OTHER INFORMATION: a, t, c or g
; FEATURE:

Query Match            5.4%;  Score 37.8; DB 4;  Length 604;
Best Local Similarity  37.5%;  Pred. No. 0.16;
Matches  102;  Conservative   0;  Mismatches 170;  Indels   0;  Gaps   0;

Qy  296  GGGGTGGAAGGAGGAGCCATGAATCAAAAGGTGGGGAAAAGGCCTTTTGGGGAGGTGG  355
              ||||| | |||||||| |||||||   |||| |||||||| |||
Db  585  GNNNANGNAGGGGGGGGGGNGGGGGGGATGGGAAANNAAGAAAAAANANANNNGGGGGGGNGG  526

Qy  356  CAGTGTGGGTAGGGAGTAGGGCTCCGGTTAGAATACATGGAAAGAAGGTGAGCACA  415
              |||||||||||  ||||| | ||||
Db  525  GGGGGGGGGGGGGGGGAAAGGGGGGGGGGAGGGGNNANAAANTANAANAAANAAAAAAA  466

Qy  416  GCCATTTCCTTTACAACAGAAATAACACAGATTCCCAGCCTCCTAACCAACAACAAG  475
                 ||||| | |||     |||       |||||
Db  465  GGAAAANTANAAAANARAATNANANNNNANGNAGGNGGAAANAAAAAAAANNN  406

Qy  476  TTTGGGAACATTCCTCCTTCTGAAATATGAAAGAGAGGGGATAAATACTGGAGTAGGAT  535
                          |||    ||||||             ||
Db  405  GGNAGNGATANNANAAANAAANAAANNNNANAAAAANAGAANAAANAANTNNAGN  346

Qy  536  TGTGAAAAAGTCAAGAGAAAAAAAAAGACA  567
                   |||| | |||||||||||||||
Db  345  NNNGNGANNNNGANAAAAAAAAAAAAAAAA  314


RESULT 4
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match            5.3%;  Score 37; DB 3;  Length 289;
Best Local Similarity  8.4%;  Pred. No. 0.19;
Matches  23;  Conservative 108;  Mismatches 142;  Indels   0;  Gaps   0;

Qy  295  GGGGGTGGGAAGGAAAAGGAGGAGCCATGAATCAAAAGGTGGGGAAAAGGCCTTTTGGGAGGTG  354
              | ::  ::| : ::|  |||: :: ::: :|:: : : :: :|
Db    5  RGRARCRARARURURARCRURARCRARARURURARCRARARURGRNRNRNRSRN  64

Qy  355  GCAGTGTGGGTAGGGAGTAGGGCTCCGGTTAGAATACATGGAAAGAAGGTGAGCAC  414
              : :: :::: : ::|  : ::| : ::::
Db   65  RNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRN  124

Qy  415  AGCCATTTCCTTTACAACAGAAATAACAGATTCCCAGCCTCCTAACCAAGAAACAACAA  474
```

Db   125 RNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRN 184
             : :   : : :   : : :   : :   : :   : : :
Qy   475 GTTTGGGAACATTCCTCTTCTTCTGAAATATGAAAGAGAGGGATAAATACTGGAGTAGGA 534
                                                |:|: : : : : : : :
Db   185 RNRSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRCRURGRCRURARARCRU 244

Qy   535 TTGTGAAAAAGTCAAGAGAAAAAAAGAACA 567
             : :: : : | | |||||||||| ||
Db   245 RCRURURGRCRGRCRUAAAAAAAAAAAAAAA 277

RESULT 5
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;   APPLICANT: Szostak, Jack W.
;   APPLICANT: Roberts, Richard W.
;   APPLICANT: Liu, Rihe
;   TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;   TITLE OF INVENTION: FUSIONS
;   FILE REFERENCE: 00786/350007
;   CURRENT APPLICATION NUMBER: US/09/244,796
;   CURRENT FILING DATE: 1999-02-05
;   EARLIER APPLICATION NUMBER: 60/035,963
;   EARLIER FILING DATE: 1997-01-27
;   EARLIER APPLICATION NUMBER: 60/064,491
;   EARLIER FILING DATE: 1997-11-06
;   EARLIER APPLICATION NUMBER: 09/007,005
;   EARLIER FILING DATE: 1998-01-14
;   NUMBER OF SEQ ID NOS: 33
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match   5.3%; Score 37; DB 3; Length 289;
Best Local Similarity   8.4%; Pred. No. 0.19; Indels   0; Gaps   0;
Matches   23; Conservative 108; Mismatches 142; Indels   0; Gaps   0;

Qy   295 GGGGGTGGGAAGGAAGAAAAGGAGGAGCCATGAATCAAAAAGTGGGGAAAAAGGCCCTTTGGGAGGTG 354
                  : :   :   :|:   :|: : : :   :|:   :   :
Db   5 RGRARCRARARURURARCRURARURURARCRARARURURARCRARARURGRNRNRNRSRN 64

Qy   355 GCAGTGTGGTAGGGAGTAGGGCTCCGGTTAGAATACATGGATGAAAGAAAGGTGAGCAC 414
             : :   : : :   :|:   : :   : : : : : :   : : :
Db   65 RNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRN 124

Qy   415 AGCCATTCCTTTACACAGAATAACAGATTTCCCAGCCTCTAACCAAAGAAGAACAACAA 474
             : :   :   :: : : : : : : : : :   : :   : : : :   :
Db   125 RNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRN 184

Qy   475 GTTTGGGAACATTCCTCTTCTTCTGAAATATGAAAGAGAGGGATAAATACTGGAGTAGGA 534
             : :   : : : :   : : : :   : : : :|:|:   : : :
Db   185 RNRSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRCRURGRCRURARARCRU 244

Qy   535 TTGTGAAAAAGTCAAGAGAAAAAAAGAACA 567
             : :: : : | | |||||||| ||
Db   245 RCRURURGRCRGRCRUAAAAAAAAAAAAAAA 277

RESULT 6
US-08-552-142A-1/c
; Sequence 1, Application US/08552142A

; Patent No. 5695995
; GENERAL INFORMATION:
;   APPLICANT: Weintraub, Harold M.
;   APPLICANT: Lee, Jacqueline E.
;   APPLICANT: Tapscott, Stephen J.
;   APPLICANT: Hollenberg, Stanley M.
;   TITLE OF INVENTION: Neurogenic Differentiation (NeuroD) Genes
;   TITLE OF INVENTION: and Proteins
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Christensen O'Connor Johnson KindnessPLLC
;   STREET: 1420 Fifth Avenue, Suite 2800
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98101-2347
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/552,142A
;   FILING DATE: 02-NOV-1995
;   CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/239,238
;   FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: WO PCT/US95/05741
;   FILING DATE: 08-MAY-1995
; ATTORNEY/AGENT INFORMATION:
;   NAME: Broderick, Thomas F.
;   REGISTRATION NUMBER: 31,332
;   REFERENCE/DOCKET NUMBER: FHCR-1-8933
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-682-8100
;   TELEFAX: 206-225-0709
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2089 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   ORIGINAL SOURCE:
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 229..1302
US-08-552-142A-1

Query Match   5.2%; Score 36.4; DB 1; Length 2089;
Best Local Similarity 52.7%; Pred. No. 0.74; Indels   0; Gaps   0;
Matches   79; Conservative   0; Mismatches 71; Indels   0; Gaps   0;

Qy   473 AAGTTTGGGAACATTCCTCCTTCTGAAATATGAAAGAGAGGGGATAAATACTGGAGTAG 532
             || ||| ||||| ||||| || || | ||| | | |||| ||||||
Db   1827 AAATTGATGAAGCTGCATGGTTTTAAAATAGAGAAATCACGTTATAAAAAGTCATTAAA 1768

Qy   533 GATTGTGAAAAAGTCAAGAGAAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTCC 592
             |||| | ||||| || | |||| |||| || || || || |||||
Db   1767 AATTCGTACAAATCACAACAAAATAATTCAGCATGGGAAAGGTCACAGGTAGTAAAAT 1708

Qy   593 ATGGGATGTAAAAGGAAGTTATTCAAT 622
             ||| | |||| ||||||| |||
Db   1707 GCTGGTTGAAAAATATATTATTATATAT 1678

RESULT 7
US-08-910-973-1/c
; Sequence 1, Application US/08910973
; Patent No. 5795723

GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson KindnessPLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/499,227
; FILING DATE: 05-August-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-May-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-May-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/910,973
; FILING DATE: 07-August-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCR-1-12742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 229..1302
US-09-499-227-1

Query Match   5.2%;   Score 36.4;   DB 4;   Length 2089;
Best Local Similarity 52.7%;   Pred. No. 0.74;
Matches   79;   Conservative   0;   Mismatches   71;   Indels   0;   Gaps   0;

Qy   473 AAGTTTGGGAACATTCCTCCTCTTCTGAAATATGAAAGAGAGGGGATAAATACTGGAGTAG 532
               || ||| | | || | | || | ||||| | | ||||| | | ||| ||| |||
Db   1827 AAAATTGATGAAGCTGCATGGTTTTAAAATAGGAAATCCACGTTTATAAAAAGTCATTAAA 1768

Qy   533 GATTGTGAAAAAGTCAAGAGAGAAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTCC 592
               || |||||| | || ||| | || | | || || ||||| || | ||
Db   1767 AATTCTGTACAAAATCACAACAAAATAATTCAGCATGGGAAAGGTCACAGGTAGTAAAAT 1708

Qy   593 ATGGGATGGTAAAAAGGAAGTTATTTCAAT 622
               | || |||||||| ||| || |||
Db   1707 GCTGGTTGAAAAATATATATTTATATATAT 1678

RESULT 9
PCT-US95-05741-1/c
; Sequence 1, Application PC/TUS9505741
; GENERAL INFORMATION:

GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson KindnessPLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,973
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCR-1-10958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 229..1302
US-08-910-973-1

Query Match   5.2%;   Score 36.4;   DB 1;   Length 2089;
Best Local Similarity 52.7%;   Pred. No. 0.74;
Matches   79;   Conservative   0;   Mismatches   71;   Indels   0;   Gaps   0;

Qy   473 AAGTTTGGGAACATTCCTCCTCTTCTGAAATATGAAAGAGAGGGGATAAATACTGGAGTAG 532
               || ||| | | || | | || | ||||| | | ||||| | | ||| ||| |||
Db   1827 AAAATTGATGAAGCTGCATGGTTTTAAAATAGGAAATCCACGTTTATAAAAAGTCATTAAA 1768

Qy   533 GATTGTGAAAAAGTCAAGAGAGAAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTCC 592
               || |||||| | || ||| | || | | || || ||||| || | ||
Db   1767 AATTCTGTACAAAATCACAACAAAATAATTCAGCATGGGAAAGGTCACAGGTAGTAAAAT 1708

Qy   593 ATGGGATGGTAAAAAGGAAGTTATTTCAAT 622
               | || |||||||| ||| || |||
Db   1707 GCTGGTTGAAAAATATATATTTATATATAT 1678

RESULT 8
US-09-499-227-1/C
; Sequence 1, Application US/09499227
; Patent No. 6444463
; GENERAL INFORMATION:

APPLICANT: Weintraub, Harold
APPLICANT: Lee, Jacqueline E.
APPLICANT: Tapscott, Stephen J.
APPLICANT: Hollenberg, Stanley M.
TITLE OF INVENTION: Neurogenic Differentiation (NeuroD) Gene
TITLE OF INVENTION: and Protein
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05741
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Broderick, Thomas F.
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: FHCR-1-8504
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-225-0709
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2089 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 229..1302
PCT-US95-05741-1

Query Match   5.2%;   Score 36.4; DB 5; Length 2089;
Best Local Similarity 52.7%; Pred. No. 0.74;
Matches   79; Conservative 0; Mismatches 71; Indels 0; Gaps   0;

Qy 473 AAGTTGGGAACATTCCTCTTCTGAAATATGAAAGAGAGGGATAAATACTGGAGTAG 532
         | |||| |||||||| ||||||
Db 1827 AAAATTGATGAAGCTGCATGGTTTAAAATAGGAAATCCACGTTATAAAAAGTCATTAAA 1768

Qy 533 GATTGTGAAAAAGTCAAGAGAAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTCC 592
         |||| ||| ||||| |||| | |||
Db 1767 AATTCTGTACAAAATCACACACAAATAATTCAGCATGGGAAAGGTCACGGTAGTAAAT 1708

Qy 593 ATGGGATGGTAAAAAGGAAGAGTTATTTCAAT 622
         |||||| |||||||| ||||| ||||
Db 1707 GCTGGTTGAAAAATATATATTTATATATAT 1678

RESULT 10
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Giordano, J.Y.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm

; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match   5.2%;   Score 36.2; DB 4; Length 832;
Best Local Similarity 11.1%; Pred. No. 0.54;
Matches   34; Conservative 147; Mismatches 125; Indels 1; Gaps   1;

Qy 159 ATATTTAAGACTGGGTAGAAAAACAAGAGAGTATGAACAAGAGTGAGGAGATTATTAGC 218
         | ||::|:: | |||||| |||:: : :: |::|
Db 310 WYMWKRMMSTRWYCYMCWKCCMYRGRRCAWYTMARGRWMSYAWGKWKSMRSAMSMCTRMY 251

Qy 219 AGTGACCTTTGAGAGTACATCTCTAGAGTGGTATGTGTAGAGCCAGATTTTCAGCATCAG 278
         :: ::|: ||| |:: |||::::||| |::|:::: :::: |::
Db 250 YKKGSTYWTMKCTCATWCYWYWKYWKRMWSKTCWSGSRGGYMTSYTSTRSYSMYWASWMY 191

Qy 279 AACCATCAAGCATTTTGGGGTGGAAGGAAGGAGCCATGAATCAAAAGGTGGGGAAAA 338
         | ::: || ||| ||||| ||| |:::: ||||||
Db 190 TMCWWWGRWWSTYWYMAWGKKWWRYATTWRRAMMWWWAAWTMWWYMWWAWCMSSRGGAAMY 131

Qy 339 GGCCTTTGGGAGGTGCAGTGTGGGTA-GGGAGTAGGGCTCCGGTTAGAATACATGGAT 397
         || :::: :::::::::|| ::::: ::::   :| :::|:
Db 130 RRTMMMWGYRYWWRKKSYRRTRCAWAYAWKTKRSYYWCWRWKWKRCMMMMMAMAYGKTM 71

Qy 398 GAAAGAAAGGTGAGCACAGCCATTTCCTTTACAACAGAATAACAGATTCCCAGCCTCT 457
         :::|:::| ::: |::|||::::|:::: ::::::|: ||
Db 70   MMRACWKTRYWRWWAWAMWRMWTMWMMYYWYWRAMKRRMWMRKWRSWSMWMAWGMTRW 11

Qy 458 AACCAAA 464
         ||:|||
Db 10   AARMWWR 4

RESULT 11
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)

us-10-051-835-14.rnl

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854)..(1664855)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match   5.2%;   Score 36.2;   DB 4;   Length 1664976;
Best Local Similarity 50.9%;   Pred No. 21;

Matches   86;   Conservative   0;   Mismatches   83;   Indels   0;   Gaps   0;

```
Qy      476  TTTGGGAACATTCCTCTTCTCTGAAATATGAAAGAGAAGAGGGGATAAATACTGGAGTAGGAT  535
             ||||||                                         |||||||
Db   751136  TTCAAGATAATCCTCTATGTTAAAATATAAATATGATAATTGAAATAAAAATGAAAC  751077

Qy      536  TGTGAAAAAGTCAAGAGAAAAAAAAGAACAGCCCAAGTGAACAGATACTTCCATG  595
             ||||||     |||||
Db   751076  AGATAAAATTTAAACAGAAAGAAAAGAGAAAAATGCATAATAATATCTTTAACTCTATATT  751017

Qy      596  GGATGGTAAAAAGGAAGTTATTCAATAAAATGACCCCTTGGAAGGAGT  644
             |||||||          |||||
Db   751016  GGAGGGCGATTAACGGAGTTATAACTTAAGGTTTCTTACTTTTTTAATGGGT  750968
```

RESULT 12
US-08-477-831C-1
; Sequence 1, Application US/08477831C
; Patent No. 6429291
; GENERAL INFORMATION:
;_ APPLICANT: TURLEY, EVA A.
;_ APPLICANT: SHUWEN, ZHANG
;_ APPLICANT: ENTWISTLE, JOYCELYN
;_ TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
;_ NUMBER OF SEQUENCES: 60
;_ CORRESPONDENCE ADDRESS:
;_   ADDRESSEE: FISH & NEAVE
;_   STREET: 1251 AVENUE OF THE AMERICAS
;_   CITY: NEW YORK
;_   STATE: NEW YORK
;_   COUNTRY: U.S.A.
;_   ZIP: 10020-1104
;_ COMPUTER READABLE FORM:
;_   MEDIUM TYPE: Floppy disk
;_   COMPUTER: IBM PC compatible
;_   OPERATING SYSTEM: PC-DOS/MS-DOS
;_   SOFTWARE: Patentin Rel. #1.0, ASCII
;_ CURRENT APPLICATION DATA:
;_   APPLICATION NUMBER: US/08/477,831C
;_   FILING DATE: 07-JUN-1995
;_   CLASSIFICATION: 514
;_ ATTORNEY/AGENT INFORMATION:
;_   NAME: PIERRI, MARGARET A.
;_   REGISTRATION NUMBER: 30,709
;_   REFERENCE/DOCKET NUMBER: SIM-10
;_ TELECOMMUNICATION INFORMATION:
;_   TELEPHONE: 212-596-9000
;_   TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 1:
;_ SEQUENCE CHARACTERISTICS:
;_   LENGTH: 1821 base pairs
;_   TYPE: nucleic acid
;_   STRANDEDNESS: single
;_   TOPOLOGY: linear
;_ MOLECULE TYPE: cDNA
;_ DESCRIPTION: /desc = "RHAMM I cDNA (coding region)"
US-08-477-831C-1

Query Match   5.2%;   Score 36;   DB 4;   Length 1821;
Best Local Similarity   45.2%;   Pred. No. 0.91;
Matches 132;   Conservative   0;   Mismatches 160;   Indels   0;   Gaps   0;

```
Qy      290  ATTTTGGGGTGGAAGGAATCAAAAGGAAAGGCCTTTTGGG  349
             ||||||                                ||||||
Db     1117  AATTTGACTCTACAAGAAAAAGTAGCTATGGCTAGGAGGCTCCGGTTAGAATAAAGTGTAGAAGATGTTCAACAG  1176

Qy      350  AGGTGGCAGTGTGGGTAGGGAGTAGGGCTCCGGTTAGAATATGGATCCTAGAGATGAAGAAGGTG  409
             ||||||
Db     1177  CAGATATTGACAGCTGAGGAGCACAAATATGCAAGGATGGTTCAGATTTGCAG  1236

Qy      410  AGCACAGCCATTCCTTTACAACAGACAGATTTCCCAGCCTCTAACCAAAGAAGGTG  469
             ||||||
Db     1237  AACAGATCAACCTTAAAGAAGAGAAATTAAAGAAATCACCATCTTCTTGAGAAAC  1296
```

```
Qy      470  AACAAGTTTGGGAACATTCCTCTTCTCTGAAATATGAAAGAGAAGAGGGGATAAATACTGGAG  529
             ||||||
Db     1297  ATAACTGATTTGAAAAATCAACTCAGACAACACAAGATGAAGACTTTAGGAAGCAGCTGGAA  1356

Qy      530  TAGGATTGTGAAAAAGTCAAGAGAAAAAAAAGAACAGCCCAAGTGTAACA  581
             |||||||
Db     1357  GGAGAAGGAAAAAGAACACAGACAGAGAAAGAAAATGTAATGACAGAGATTAACCA  1408
```

RESULT 13
US-08-477-831C-9
; Sequence 9, Application US/08477831C
; Patent No. 6429291
; GENERAL INFORMATION:
;_ APPLICANT: TURLEY, EVA A.
;_ APPLICANT: SHUWEN, ZHANG
;_ APPLICANT: ENTWISTLE, JOYCELYN
;_ TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
;_ NUMBER OF SEQUENCES: 60
;_ CORRESPONDENCE ADDRESS:
;_   ADDRESSEE: FISH & NEAVE
;_   STREET: 1251 AVENUE OF THE AMERICAS
;_   CITY: NEW YORK
;_   STATE: NEW YORK
;_   COUNTRY: U.S.A.
;_   ZIP: 10020-1104
;_ COMPUTER READABLE FORM:
;_   MEDIUM TYPE: Floppy disk
;_   COMPUTER: IBM PC compatible
;_   OPERATING SYSTEM: PC-DOS/MS-DOS
;_   SOFTWARE: Patentin Rel. #1.0, ASCII
;_ CURRENT APPLICATION DATA:
;_   APPLICATION NUMBER: US/08/477,831C
;_   FILING DATE: 07-JUN-1995
;_   CLASSIFICATION: 514
;_ ATTORNEY/AGENT INFORMATION:
;_   NAME: PIERRI, MARGARET A.
;_   REGISTRATION NUMBER: 30,709
;_   REFERENCE/DOCKET NUMBER: SIM-10
;_ TELECOMMUNICATION INFORMATION:
;_   TELEPHONE: 212-596-9000
;_   TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 9:
;_ SEQUENCE CHARACTERISTICS:
;_   LENGTH: 1885 base pairs
;_   TYPE: nucleic acid
;_   STRANDEDNESS: single
;_   TOPOLOGY: linear
;_ MOLECULE TYPE: cDNA
;_ DESCRIPTION: /desc = "RHAMM IB"
US-08-477-831C-9

Query Match   5.2%;   Score 36;   DB 4;   Length 1885;
Best Local Similarity   45.2%;   Pred. No. 0.92;
Matches 132;   Conservative   0;   Mismatches 160;   Indels   0;   Gaps   0;

```
Qy      290  ATTTTGGGGTGGAAGGAAAAGGAGCCATGAATCAAAAGGTGGGGAAAAGGCCTTTTGGG  349
             ||||||                                ||||||
Db     1181  AATTTGACTCTACAAGAAAAAGTAGCTTATGGCTGGCTAGGAGGCTCCGGTTAGAATACATGGATGAAGAAGGTG  1240

Qy      350  AGGTGGCAGTGTGGGTAGGGAGTAGGGCTCCGGTTAGAATATGCAAGATATGCAAGGATGGTTCAGATTTGCAG  1300
             ||||||
Db     1241  CAGATATTGACAGCTGAGGAGCACAAATCAAGAAATATAACAGATTTCCCAGCCTCTAACCAAAGAAAC  469

Qy      410  AGCACAGCCATTCCTTTACAACAGACAGATTTCCCAGCCTCTAACCAAAGAAAC  469
             ||||||
Db     1301  AACAGATCAACCTTAAAGAAGAGAAATTAAAGAAATCACATCTTCTTGAAATATGAAAGAGAGGGGATAAATACTGGAG  1360

Qy      470  AACAAGTTTGGGAACATTCCTCTTCTCTGAAATATGAAAGAGAGGGGATAAATACTGGAG  529
             ||||||
Db     1361  ATAACTGATTTGAAAAATCAACTCAGACAACAGATGAAGACTTTAGGAAGCAGCTGGAA  1420
```

```
Qy     530  TAGGATTGTGAAAAAAGTCAAGAGAGAAAAAAAGAGACAGCCCAGTGTAACA  581
                 ||||||||      |||||||    ||||      ||||||||
Db    1421  GAGAAAGGAAAAAAGACAGCAGAGAGAAAGAAGAAAATGTAATGACAGAATTAACCA 1472
```

RESULT 15
US-08-477-831C-8
; Sequence 8, Application US/08477831C
; Patent No. 6429291
  GENERAL INFORMATION:
;   APPLICANT: TURLEY, EVA A.
;   APPLICANT: SHUWEN, ZHANG
;   APPLICANT: ENTWISTLE, JOYCELYN
;   TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
;   NUMBER OF SEQUENCES: 60
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FISH & NEAVE
;   STREET: 1251 AVENUE OF THE AMERICAS
;   CITY: NEW YORK
;   STATE: NEW YORK
;   COUNTRY: U.S.A.
;   ZIP: 10020-1104
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Rel. #1.0, ASCII
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/477,831C
;   FILING DATE: 07-JUN-1995
;   CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
;   NAME: PIERRI, MARGARET A.
;   REGISTRATION NUMBER: 30,709
;   REFERENCE/DOCKET NUMBER: SIM-10
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 212-596-9000
;   TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1961 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;   DESCRIPTION: /desc = "RHAMM IA"
US-08-477-831C-8

Query Match   5.2%; Score 36; DB 4; Length 1961;
Best Local Similarity 45.2%; Pred. No. 0.94;
Matches 132; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy   290 ATTTTGGGGTGGAAGGAAAAAGGAGCCATGAATCAAAAGGTGGGAAAAGGCCTTTTGGG 349
                 |||||| |||||| |||||| |||||||
Db   1257 AATTTGACTTCTACAAGAAAAAAGTAGCTATGGCTATGGCTGAAAAAAAGTGTAGAGATGTTCAACAG 1316

Qy   350 AGGTGGCAGTGTGGGTAGGAGTAGGGCTCCGGTTAGAATACATGGATGGTTCAGCAG 409
                 |||||| |||||| ||||||
Db   1317 CAGATATTGACAGCTGAGAGCACAAATCAAGAATATGCAAGGATGGTTCAGGATTTGCAG 1376

Qy   410 AGCACAGCCATTCCTTACCTAAAAGAAGAAGAATTAAAGAAATAACAGATTTCCCAGCCTCTAACCAAAGAAAC 469
                 |||||| ||||||
Db   1377 AACAGATCAACCTTAAAAGAAGAAGAAATTAAAGAAATCACATCTTCATTTCTTGAGAAA 1436

Qy   470 AACAAGTTTGGGAACATTCCTCCTTCTTGAAAATATGAAAGAAGAGGGGATAAATAAACTGGAG 529
                 |||||| |||||||
Db   1437 ATAACTGATTTGAAAAATCAACTCAGACAACAAGAATGAAGACAGCCCAAGTGTAACA 1496

Qy   530 TAGGATTGTGAAAAAAGTCAAGAGAGAAAAAAAGAACAGCCCAGTGTAACA 581
                 |||||||| |||||| |||| |||||||
Db   1497 GAGAAAGGAAAAAAGAACAGCGAGAAAATGTAATGACAGAATTAACCA 1548
```

Search completed: April 26, 2004, 02:11:00
Job time : 40.5277 secs

RESULT 14
US-08-477-831C-10
; Sequence 10, Application US/08477831C
; Patent No. 6429291
  GENERAL INFORMATION:
;   APPLICANT: TURLEY, EVA A.
;   APPLICANT: SHUWEN, ZHANG
;   APPLICANT: ENTWISTLE, JOYCELYN
;   TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
;   NUMBER OF SEQUENCES: 60
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FISH & NEAVE
;   STREET: 1251 AVENUE OF THE AMERICAS
;   CITY: NEW YORK
;   STATE: NEW YORK
;   COUNTRY: U.S.A.
;   ZIP: 10020-1104
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Rel. #1.0, ASCII
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/477,831C
;   FILING DATE: 07-JUN-1995
;   CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
;   NAME: PIERRI, MARGARET A.
;   REGISTRATION NUMBER: 30,709
;   REFERENCE/DOCKET NUMBER: SIM-10
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 212-596-9000
;   TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1896 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;   DESCRIPTION: /desc = "RHAMM I-2a"
US-08-477-831C-10

Query Match   5.2%; Score 36; DB 4; Length 1896;
Best Local Similarity 45.2%; Pred. No. 0.93;
Matches 132; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

```
Qy     290  ATTTTGGGGTGGAAGGAAAAAGGAGCCATGAATCAAAAGGTGGGAAAAGGCCTTTTGGG  349
                 ||||||      ||||||      ||||||
Db    1192  AATTTGACTTCTACAAGAAAAAAGTAGCTATGGCTATGGCTGAAAAAAAGTGTAGAGATGTTCAACAG 1251

Qy     350  AGGTGGCAGTGTGGGTAGGAGTAGGGCTCCGGTTAGAATACATGGATGGTTCAGCAG  409
                 ||||||
Db    1252  CAGATATTGACAGCTGAGAGCACAAATCAAGAATATGCAAGGATGGTTCAGGATTTGCAG 1311

Qy     410  AGCACAGCCATTCCTTACACAGAAGAAGAATTCCCAGCCTCTAACCAAAGAAAC  469
                 ||||||
Db    1312  AACAGATCAACCTTAAAAGAAGAAGAATTAAAGAAATCACTCTTCATTTCTTGAGAAA 1371

Qy     470  AACAAGTTTGGGAACATTCCTCCTTCTGAAAATATGAAAGAGGGGATAAATACTGGAG  529
                 ||||||      |||||||
Db    1372  ATAACTGATTTGAAAAATCAACTCAGACACACAAGAATGAAGACTTTAGGAGCAGCTGGAA 1431

Qy     530  TAGGATTGTGAAAAAAGTCAAGAGAGAAAAAAAGAACAGCCCAGTGTAACA  581
                 ||||||||      ||||||    ||||      |||||||
Db    1432  GAGAAAGGAAAAAAGACAGCAGAGAGAAAGAAGAAAATGTAATGACAGAATTAACCA 1483
```

```
                GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    April 24, 2004, 23:47:42 ; Search time 184.93 Seconds
                                       (without alignments)
                                       15988.422 Million cell updates/sec

Title:            US-10-051-835-14
Perfect score:    696
Sequence:         1 taaaatatctgataggcagt.........tttgaaaatatataaaat 696

Scoring table:    IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched:         3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:     6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  N_Geneseq_29Jan04:*
            1: geneseqn1980s:*
            2: geneseqn1990s:*
            3: geneseqn2000s:*
            4: geneseqn2001as:*
            5: geneseqn2001bs:*
            6: geneseqn2002s:*
            7: geneseqn2003as:*
            8: geneseqn2003bs:*
            9: geneseqn2003cs:*
           10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
c 1	681.8	98.0	17947	4	AAK77679	Aak77679 Human imm
c 2	44.8	6.4	9415	6	ABL32295	Abl32295 Human imm
3	43.4	6.3	2000	7	ADA71938	Ada71938 Rice gene
4	41.8	6.0	300	4	AAI23067	Aai23067 Probe #13
5	41.8	6.0	300	4	ABA68161	Aba68161 Human foe
6	41.8	6.0	300	4	AAI48373	Aai48373 Probe #17
7	41.8	6.0	300	4	ABA50219	Aba50219 Human bre
8	41.8	6.0	300	4	AAK16541	Aak16541 Human bra
9	41.8	6.0	300	6	ABS41903	Abs41903 Human liv
10	41.8	6.0	300	6	ABS16349	Abs16349 Human gen
11	41	5.9	451	6	ABN64891	Abn64891 Human can
c 12	40.8	5.9	6305	2	AAQ96127	Aaq96127 Tobacco 1
c 13	40.6	5.8	6668	6	ABL33696	Abl33696 Human imm
14	40.2	5.8	1269	7	ACA28673	Aca28673 Prokaryot
15	39.8	5.7	32249	5	ABA17155	Aba17155 Human ner
16	39.6	5.7	6319	6	ABL33643	Abl33643 Human imm
17	39.6	5.7	83391	6	ABQ67094	Abq67094 Human ang
18	39.4	5.7	6681	6	ABL32155	Abl32155 Human imm
19	39.4	5.7	6681	6	ABL54304	Abl54304 Chemical
20	39.4	5.7	46358	7	ACC42503	Acc42503 Murine P2
21	39	5.6	9721	6	AAK65240	Aak65240 Human imm
c 22	38.8	5.6	10048	6	ABL70314	Abl70314 Chemical
c 23	38.8	5.6	10048	6	AAS61252	Aas61252 Human gen
24	38.4	5.5	421	6	ABV98460	Abv98460 Human pan
25	38.4	5.5	657	4	ABA08927	Aba08927 Human nov
26	38.4	5.5	699	4	AAL00184	Aal00184 Human rep
27	38.4	5.5	857	9	ADD68343	Add68343 Chicken s
28	38.4	5.5	1079	6	ABZ20626	Abz20626 Human cyt
29	38.4	5.5	1686	6	ABZ11445	Abz11445 Human pol
30	38.4	5.5	1804	5	AAH15366	Aah15366 Human cDN
c 31	38.4	5.5	2291	6	ABX71418	Abx71418 Cell stru
c 32	38.4	5.5	14861	6	ABL33194	Abl33194 Human imm
c 33	38.4	5.5	14861	6	ABL70553	Abl70553 Chemicall
c 34	38	5.5	14861	6	AAS61202	Aas61202 Human gen
c 35	38	5.5	5823	6	ABK28382	Abk28382 DNA trans
c 36	37.8	5.4	5660	6	ABL34357	Abl34357 Human imm
37	37.6	5.4	723	6	ABQ25179	Abq25179 Oligonucl
38	37.6	5.4	723	6	ABQ25178	Abq25178 Oligonucl
c 39	37.6	5.4	6136	6	ABK40042	Abk40042 Human che
40	37.6	5.4	9539	6	AAS45346	Aas45346 Chemical
c 41	37.6	5.4	9539	6	ABK28179	Abk28179 DNA trans
c 42	37.6	5.4	14798	6	ABL33033	Abl33033 Human imm
43	37.6	5.4	96593	8	ADA02561	Ada02561 Mouse Ras
44	37.6	5.4	96593	9	ADB72299	Adb72299 Mouse Ras
45	37.4	5.4	410	6	ABV98193	Abv98193 Human pan

ALIGNMENTS

```
RESULT 1
AAK77679/c
ID   AAK77679 standard; DNA; 17947 BP.
XX
AC   AAK77679;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32491.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214486P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
```

```
PR   14-AUG-2000;   2000US-0252270P.
PR   14-AUG-2000;   2000US-0254447P.
PR   14-AUG-2000;   2000US-0257557P.
PR   14-AUG-2000;   2000US-0257558P.
PR   14-AUG-2000;   2000US-0257559P.
PR   18-AUG-2000;   2000US-0262799P.
PR   22-AUG-2000;   2000US-0266681P.
PR   22-AUG-2000;   2000US-0266868P.
PR   22-AUG-2000;   2000US-0271182P.
PR   23-AUG-2000;   2000US-0270009P.
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0292287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   05-SEP-2000;   2000US-0229345P.
PR   06-SEP-2000;   2000US-0230437P.
PR   06-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0244474P.
PR   08-NOV-2000;   2000US-0244475P.
PR   08-NOV-2000;   2000US-0244476P.
PR   08-NOV-2000;   2000US-0244477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0245532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251889P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 32491; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastase of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 17947 BP; 5405 A; 3428 C; 3406 G; 5708 T; 0 U; 0 Other;

Query Match             98.0%;  Score 681.8;  DB 4;  Length 17947;
Best Local Similarity   99.6%;  Pred. No. 2.6e-177;
Matches  694;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    1;
```

RESULT 2
ABL32295/c
ID   ABL32295 standard; DNA; 9415 BP.
XX
AC   ABL32295;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 268.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX

PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 268; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 9415 BP; 1906 A; 297 C; 2611 G; 4601 T; 0 U; 0 Other;

Query Match   6.4%; Score 44.8; DB 6; Length 9415;
Best Local Similarity 49.2%; Pred. No. 0.1;
Matches 118; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Qy   1 TAAAATATCTGATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGAGGAGGCTTTGATG 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10497 TAAAATATCAAATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGAGGAGGCTTTGATG 10438

Qy   61 GCGATACAGACTTGGAAGACATCAGTGCTGAGCAGTAAATGAGATGATTCAGGAAAGAGT 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10437 GCGATACAGACTTGGAAGACATCAGTGCTGAGCAGTAAATGAGATGATTCAGGAAAGAGT 10378

Qy   121 ATAAACTGGGAAGAGGACAGAGGACAGGCTCAAGGAACATATTTAAGGACTGGGTAGAAA 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10377 ATAAACTGGGAAGAGGACAGAGGACAGGCTCAAGGAACATATTTAAGGACTGGGTAGAAA 10318

Qy   181 AACAAGAGAGTATGAACAAGAGTGAGGAGATTATTAGCAGTGACCTTTGAGAGTACATCT 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10317 AACAAGAGAGTATGAACAAGAGTGAGGAGATTATTAGCAGTGACCTTTGAGAGTACATCT 10258

Qy   241 CTAGAGTCGTATGTGTAGAGCCAGATTTCAGCATCAGAACCATCAAGCATTTTGGGGGT 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10257 CTAGAGTCGTATGTGTAGAGCCAGATTTCAGCATCAGAACCATCAAGCATTTTGGGGGT 10198

Qy   301 GGAAGGAAAAGGAGCCATGAATCAAAAGTGGGGAAAAGGCCTTTGGGAGTGGCAGTG 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10197 GGAAGGAAAAGGAGCCATGAATCAAAAGTGGGGAAAAGGCCTTTGGGAGTGGCAGTG 10138

Qy   361 TGGGTAGGGAGTAGGGCTCCGGTTAGAATACATGGATGAAAGAAAGTGAGCACAGCCAT 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10137 TGGGTAGGGAGTAGGGCTCCGGTTAGAATACATGGATGAAAGAAAGTGAGCACAGCCAT 10078

Qy   421 TTCCTTTACAACAGAAATAACAGATTTCCAGCCTCTAACCAAAGAAACAACAAGTTTGG 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10077 TTCCTTTACAACAGAAATAACAGATTTCCAGCCTCTAACCAAAGAAACAACAAGTTTGG 10018

Qy   481 GAACATTCCTCCTCTTCTGAAATATGAAAAGAGAGAGGGGGATAAAATACTGGAGTAGGATTGTGA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10017 GAACATTCCTCCTCTTCTGAAATATGAAAAGAGAGAGGGGGATAAAATACTGGAGTAGGATTGTGA 9958

Qy   541 AAAAAGTCAAGAG-AAAAAAAAAAAGAAAAAAGAACAGCCCAAGTGTAACAACAACAAGTT 599
            |||||||||||||| ||||||||||||||||||||||||||||||
Db   10017 AAAAAGTCAAGAGAAAAAAAAAAAAGAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTCATGGGAT 9898

Qy   600 GGTAAAAGGGAAGTTATTTCAATAAAATGACCCCTTGGAAGGAGTTCAAAACAGGTTGCC 659
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9897 GGTAAAAGGGAAGTTATTTCAATAAAATGACCCCTTGGAAGGAGTTCAAAACAGGTTGCC 9838

Qy   660 ATATGCTTTATGTGGAGTTTTGAAAAAATATATAAAAT 696
            ||||||||||||||||||||||||||||||||||||||
Db   9837 ATATGCTTTATGTGGAGTTTTGAAAAAATATATAAAAT 9801

RESULT 3
ADA71938
ID   ADA71938 standard; DNA; 2000 BP.
XX
AC   ADA71938;
XX
DT   20-NOV-2003 (first entry)
XX
DE   Rice gene, SEQ ID 5263.
XX
KW   Plant; bacterial infection; fungal infection; viral infection; rice;
KW   gene; ds.
XX
OS   Oryza sativa.
XX
PN   WO2003000898-A1.
XX
PD   03-JAN-2003.
XX
PF   22-JUN-2001; 2001WO-IB001105.
XX
PR   22-JUN-2001; 2001WO-IB001105.
XX

Qy   387 AATACACATGGATGAAAGAAAGGTGAGCACAGCCATTTCCTTTACACAGAAATAACAGATT 446
            |||||||||||||||||||||||||||||||
Db   7617 AATAACCGTAAAAAAATATAAAA0CGTAAAAAAATTTCCCACTAAAAAAAAAATTTAAAAAAAA 7558

Qy   447 TCCCAGCCTCTAACCAAAGAAACAACAACAAGTTGGGAACATTCCTCCTTCTGAAATATGA 506
            |||||||||||||||||||||||||||||||
Db   7557 ACTCCGTCCAAACTAAAAAAAAAAAACGCGTTAAAAAAAAGTCAACCGGATAAAAAAAAAC 7498

Qy   507 AAGAGAGGGGGATAAAATACTGGAGTAGGAGGATTGTGAAAAAAGTCAAGAGAAAAAAGAAC 566
            ||||||||||||||||||||||||| ||
Db   7497 AAAAAACCGAAAAATAAAACCAAAATAAAAAAAAACACCGAATTAAAAAAAAAAAAAAAA 7438

Qy   567 AGCCCAAGTGTAACAGATACTTCTCCATGGGATGGTAAAAGGAAGTTATTTCAATAAAA 626
            |||||||||||||||||||||||||
Db   7437 CGACCAAAACTAACAACTTCCTAAATCTAAAATTAAATAAAAAAAAAAAAACTAACTAAAA 7378

```
XX
PA  (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI  Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI  Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR  WPI; 2003-175290/17.
XX
PT  Identifying at least one gene involved in plant resistance or response to
PT  pathogenic infection for conferring resistance or tolerance to a plant to
PT  bacterial, fungal or viral infection by determining or detecting plant
PT  gene expression.
XX
PS  Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC  The present invention relates to a method (M1) for identifying genes
CC  involved in plant resistance or response to pathogenic infection. M1
CC  comprises identifying a gene whose expression is significantly altered in
CC  the incompatible interaction of plant gene expression relative to
CC  expression of the gene in an uninfected plant, in a mutant plant that
CC  does not express a gene associated with response to pathogenic infection,
CC  or in a corresponding incompatible or compatible interaction. (M1) is
CC  useful for conferring resistance to resistance or tolerance to a plant to
CC  bacterial, fungal or viral infection. The present sequence was used to
CC  illustrate the invention.
XX
SQ  Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
```

Query Match   6.3%; Score 43.8; DB 7; Length 2000;
Best Local Similarity   9.9%; Pred. No. 0.11;
Matches   69; Conservative 292; Mismatches 334; Indels 0; Gaps 0;

```
Qy      2 AAAATATCTGATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGGAGGAGAGGCTTTGATGG 61
           :: :||::: ||:|:|||| ::: :|| ::|| ::  |::|:
Db    307 WRGATRYWGRGYMSRMAMMYKKMYWYRGYKGMKRGWWAGRMMMRSMCRKSKACYMRWRM 366

Qy     62 CGATACAGACTTGGAAGACATCAGTGCTGAGCAGTAAATGAGATGATTCAGGAAAGAGTA 121
           ::|: ||:: ::||| ||:  :| :|:  :|:|:: ||::|: :|:|
Db    367 WRMTRRRRWAKKSSRTSRRKRKWCMRKRKYKRWRGYSRMRSCKRARWMKRCRSGRAWKM 426

Qy    122 TAAACTGGGAAGGAGAGGACAGGAGGACCAGGCTCAAGGAACATATTTAAGGACTGGGTAGAAAA 181
           :| :  |:|::::|:|  : |::|:::| :|:|||  ::|| :: :|:
Db    427 GCRGCMTCRMKSYGMWRWKSWKRMASKYKWMSRMYRWRKKKCSRTTMWGKTRGGWMGTMG 486

Qy    182 ACAAGAGAGTATGAACAAGAGTGAGGAGATTATTAGCAGTGACCTTGAGAGTACATCTC 241
           ||  |:|:  ::| :|::||  ||:: ::|:
Db    487 RCRYKKRSGMKRKCRRRRWGRMYRMWRKRYMSARYTMRYCARKKYSYSAARKARCWYRG 546

Qy    242 TAGAGTGGTATGTGTAGAGCCAGATTTCAGCATCAGAACCATCAAGCATTTGGGGGTG 301
           :|: :: : :||: :| |:: :| ||::|: :|: :|:
Db    547 KGYYWAGMWWKRYKRNYMYKMWMWYRKRYKYSKCSWYCKMSYYASCMKSARKAGAKMCKRSK 606

Qy    302 GAAGGAAAAGGAGCCATGAATCAAAAGGTGGGAAAAAGGCCTTTTGGGAGGTGGCAGTGT 361
           :| ||:|||:| ||| :: ||::: :|::: ::||:|:| :|
Db    607 MSAWSKSMSRSRKCRKCASKRSSAKRYANMGGMTSGSRMSRWKSYTCYWRKWGSMKSTCT 666

Qy    362 GGGTAGGGAGTAGGGCTCCGGTTAGAATACATGGATGAAAGAAAGGTGAGCACAGCCATT 421
           ::|:  :|  ||:|:: |:|:| ::: ::|: ::|:||:|:|
Db    667 WMYYMSKYTYAKYGSYWRYRYRAWCMYMWRWYYYRYRSYMTYMAWYTSSTRMAMTGMKYS 726

Qy    422 TCCTTTACAACAGAAATAACADAGATTTCCCAGCCTTCTAACCAAAGAAACAACAAGTTTGGG 481
           :::: | |:|| |: || : : |:|: : ||  :||| ::::| :|
Db    727 GRYWTSWYKYCKCSWKYRSRMWYYWSWWWAKTWKMWRRYATRMMWMWYRYSMKWYTWCTMW 786

Qy    482 AACATTCCTCTCTTCTGAAATATGAAAGAGAGGGGATAAATACTGGAGTAGGATTGTGAA 541
            ::| ::|  | |: ||:
Db    787 GYWWYWWRTYMCQRYNYMYKCTKTYWYWSATYWTGTWAAWWMAKTKMRMGMTGAKTRGRAR 846

Qy    542 AAAAGTCAAGAGAAAAAAAAAAAGAACACAGCCCAAGTGTAACACAGATACTTCTCCATGGGATGG 601
           |: |:| | :::  :|: || :|: :::: :| :|:|
Db    847 KARYWHKWATWCATKRWMTKGGAKWAWTWMAKAWRKYYWSMRRAWYYYYKTRRTRYKTC 906

Qy    602 TAAAAAGGAAGTTATTTCAATAAAATGACCCCCTTGGAAGGAGTTCAAAACAGGTTGCCAT 661
```

```
Db    907 WWKARWGSWAYRWRMWKGSAKMWMWKGGRWGWTKYWYWYCTTWKMACGRATKYMCCAGW 966

Qy    662 ATGCTTTATGTGGAGTTTGAAAAATATATAAAT 696

Db    967 WAMYSYSWTRTYMRTWRWMWASSRTAKRMARMMW 1001
```

RESULT 4
AAI23067/c
ID   AAI23067 standard; DNA; 300 BP.
XX
AC   AAI23067;
XX
DT   12-OCT-2001 (first entry)
XX
DE   Probe #13000 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 25; SEQ ID NO 13000; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging of
CC   diseases of the cervix, notably cervical cancer. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 300 BP; 68 A; 87 C; 37 G; 108 T; 0 U; 0 Other;

Query Match   6.0%; Score 41.8; DB 4; Length 300;
Best Local Similarity 52.5%; Pred. No. 0.19;
Matches 115; Conservative   0; Mismatches 102; Indels 2; Gaps 1;

```
Qy      3 AAATATCTGATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGGAGGAGGGCTTTGATGGC 62
           |||||   |||||  |||||| |   |||||  ||||||||||   |||||||||| ||
Db    228 AAATGTGAAGTTTGCAGTTGAATATTTGTCCTTGGAACATTAGGAGAAAAACAATGTGTGGGC 169

Qy     63 GA--TACAGACTTGGAAGACATCAGTGCTGAGCAGTAAATGAGATGATTCAGGAAAAGAGT 120
           ||  ||||| |||   |||| || ||  |    |||   |||  ||||||  ||||||||||
Db    168 CATCTGCATATCGGTGATGAATGAAGAAGCAGCCACAGAGCAGAAGCAGCAATCAATGCAGCAAGAGT 109
```

Qy 121 ATAAACTGGGAAGAGAGGACAGAGGACAGGCTCAAGGAACATATTTAAGGACTGGGTAGAAA 180
        |||||||| |||||| ||||||||
Db 108 AAATAAAGTGAAGAGAGTGAACAAGAGTGAGGAAGAGAGAAGAAGCCAATGTTTAAGGG 49

Qy 181 AACAAGAGAGTATGAACAAGAGTGAGGAGATTATTAGCA 219
        |||||| |||| || ||||
Db   48 AATGATAGTGTATTCAAAATTGGCAAAAGAGACTGAGAA 10

RESULT 5
ABA68161/c
ID   ABA68161 standard; DNA; 300 BP.
XX
AC   ABA68161;
XX
DT   01-FEB-2002 (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #16466.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 4; SEQ ID NO 16466; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 300 BP; 68 A; 87 C; 37 G; 108 T; 0 U; 0 Other;

Query Match   6.0%; Score 41.8; DB 4; Length 300;
Best Local Similarity   52.5%; Pred. No. 0.19;
Matches 115; Conservative   0; Mismatches 102; Indels   2; Gaps   1;

Qy   3 AAATATCTGATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGAGAGGCTTTGATGGC 62
        |||| |||| |||||| ||| ||||
Db 228 AAATGTGAAGTTTGCAGTTGAATATTTGTGCCTGGAACTTAGGAAAAACAATGTGTGGGC 169

Qy   63 GA--TACAGACTTGGAAGACATCAGTGCTGAGCAGTAAATGAGATGATTCACGAAAGAGT 120
        || |||| ||| ||| ||||
Db 168 CATCTGCGATATCGGTGATGAATGAGCAAGCCACAGAAGCAGAAGCCAATCAATGCAGCAAGAGT 109

Qy 121 ATAAACTGGGAAGAGAGGACAGAGGACAGGCTCAAGGAACATATTTAAGGACTGGGTAGAAA 180

Db 108 AAATAAAGTGAAGAGAGTGAACAAGAGTGAGGAAGAGAGAAGAAGCCAATGTTTAAGGG 49

Qy 181 AACAAGAGTATGAACAAGAGTGAGGAGATTATTAGCA 219

Db   48 AATGATAGTGTATTCAAAATTGGCAAAAGAGACTGAGAA 10

RESULT 6
AAI48373/c
ID   AAI48373 standard; DNA; 300 BP.
XX
AC   AAI48373;
XX
DT   17-OCT-2001 (first entry)
XX
DE   Probe #17059 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000663.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human placenta.
XX
PS   Claim 25; SEQ ID NO 17059; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders
XX
SQ   Sequence 300 BP; 68 A; 87 C; 37 G; 108 T; 0 U; 0 Other;

Query Match   6.0%; Score 41.8; DB 4; Length 300;
Best Local Similarity   52.5%; Pred. No. 0.19;
Matches 115; Conservative   0; Mismatches 102; Indels   2; Gaps   1;

Qy   3 AAATATCTGATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGAGAGGCTTTGATGGC 62
        |||| |||| |||||| ||| ||||
Db 228 AAATGTGAAGTTTGCAGTTGAATATTTGTGCCTGGAACTTAGGAAAAACAATGTGTGGGC 169

Qy   63 GA--TACAGACTTGGAAGACATCAGTGCTGAGCAGTAAATGAGATGATTCAATCGCAGCAAGAGT 120
        || |||| ||| ||| ||||
Db 168 CATCTGCGATATCGGTGATGAATGAGCAAGCCACAGAAGCAGAAGCCAATCAATGCAGCAAGAGT 109

Qy 121 ATAAACTGGGAAGAGAGGACAGAGGACAGGCTCAAGGAACATATTTAAGGACTGGGTAGAAA 180

Db 108 AAATAAAGTGAAGAGAGTGAACAAGAGTGAGGAAGAGAGAAGAAGCCAATGTTTAAGGG 49

Qy 181 AACAAGAGTATGAACAAGAGTGAGGAGATTATTAGCA 219

```
Db   48 AATGATAGTGTATTCAAAATTGGCAAAAGAGACTGAGAA 10
```

RESULT 7
ABA50219/c
```
ID  ABA50219 standard; DNA; 300 BP.
XX
AC  ABA50219;
XX
DT  01-FEB-2002 (first entry)
XX
DE  Human breast cell single exon nucleic acid probe #8914.
XX
KW  Human; microarray; single exon probe; gene expression; breast; disease;
KW  cancer; ss.
XX
OS  Homo sapiens.
XX
PN  WO200157271-A2.
XX
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US000662.
XX
PR  04-FEB-2000; 2000US-0180312P.
PR  26-MAY-2000; 2000US-0207456P.
PR  30-JUN-2000; 2000US-00608408.
PR  03-AUG-2000; 2000US-00632366.
PR  21-SEP-2000; 2000US-0234687P.
PR  27-SEP-2000; 2000US-0236359P.
PR  04-OCT-2000; 2000GB-00024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR  WPI; 2001-496933/54.
XX
PT  New spatially-addressable set of single exon nucleic acid probes, useful
PT  for measuring gene expression in sample derived from human breast,
PT  comprises number of single exon nucleic acid probes.
XX
PS  Claim 4; SEQ ID NO 8914; 327pp + Sequence Listing; English.
XX
CC  The invention relates to a spatially-addressable set of single exon
CC  nucleic acid probes for measuring gene expression in a sample derived
CC  from human breast and BT 474 cells. The method involves contacting the
CC  probes with a collection of detectably labelled nucleic acids derived
CC  from mRNA of human breast, and then measuring the label bound to each
CC  probe of the microarray. The probes are useful for verifying the
CC  expression of regions of genomic DNA predicted to encode proteins. They
CC  are useful for gene discovery, and for determining predisposition and/or
CC  prognosing breast disease. Gene expression analysis is useful for
CC  assessing the toxicity of chemical agents on cells. The microarray of
CC  this invention presents a far greater diversity of probes for measuring
CC  gene expression, with far less bias than expressed sequence tag
CC  microarrays. The method is suitable for rapid production of functional
CC  information from genomic sequence. The present sequence is a single exon
CC  nucleic acid probe of the invention. Note: The sequence data for this
CC  patent did not form part of the printed specification, but was obtained
CC  in electronic format directly from WIPO at
CC  ftp.wipo.int/pub/published_pct_sequences
XX
SQ  Sequence 300 BP; 68 A; 87 C; 37 G; 108 T; 0 U; 0 Other;
```

Query Match   6.0%; Score 41.8; DB 4; Length 300;
Best Local Similarity 52.5%; Pred. No. 0.19;
Matches 115; Conservative 0; Mismatches 102; Indels 2; Gaps 1;

```
Qy    3 AAATATCTGATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGAGAGGGCTTTGATGGC 62
Db  228 AAATGTGAAGTTTGCAGTTGAATATTTGTGCCTGGAACTTAGGAAAAACAATGTGTGGGC 169
Qy   63 GA--TACAGACTTGGAAGACATCAGTCGTGAGCAGTAAATGAGATGATTCAGGAAAGAGT 120
Db  168 CATCTGCATATCGGTGATGAATGAAGCCACAGAAGCAGAAGCCATCATGCAGCAAGAGT 109
Qy  121 ATAAACTGGGAAGAGACAGAGGACAGAGCTCAAGGAACATATTTAAGGACTGGGTAGAAA 180
Db  108 AATAAAGTGAAGAGGGAAGAGGAGGAGAAAGAGAGAAGAGAAGCCAATGTTTAAGGG 49
Qy  181 AACAAGAGAGTATGAACAGAGTGAGGAGGATTATTAGCA 219
Db   48 AATGATAGTGTATTCAAAATTGGCAAAAGAGACTGAGAA 10
```

RESULT 8
AAK16541/c
```
ID  AAK16541 standard; DNA; 300 BP.
XX
AC  AAK16541;
XX
DT  05-NOV-2001 (first entry)
XX
DE  Human brain expressed single exon probe SEQ ID NO: 16532.
XX
KW  Human; brain expressed exon; gene expression analysis; probe; microarray;
KW  Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW  ss.
XX
OS  Homo sapiens.
XX
PN  WO200157275-A2.
XX
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US000667.
XX
PR  04-FEB-2000; 2000US-0180312P.
PR  26-MAY-2000; 2000US-0207456P.
PR  30-JUN-2000; 2000US-00608408.
PR  03-AUG-2000; 2000US-00632366.
PR  21-SEP-2000; 2000US-0234687P.
PR  27-SEP-2000; 2000US-0236359P.
PR  04-OCT-2000; 2000GB-00024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR  WPI; 2001-483446/52.
XX
PT  Single exon nucleic acid probes for analyzing gene expression in human
PT  brains.
XX
PS  Example 4; SEQ ID NO 16532; 650pp + Sequence Listing; English.
XX
CC  The present invention provides a number of single exon nucleic acid
CC  probes which are derived from genomic sequences expressed in the human
CC  brain. They can be used to measure gene expression in brain cell samples,
CC  which may enable the diagnosis and improved treatment of nervous system
CC  diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC  epilepsy and cancers. The present sequence is one of the probes of the
CC  invention
XX
SQ  Sequence 300 BP; 68 A; 87 C; 37 G; 108 T; 0 U; 0 Other;
```

Query Match   6.0%; Score 41.8; DB 4; Length 300;
Best Local Similarity 52.5%; Pred. No. 0.19;
Matches 115; Conservative 0; Mismatches 102; Indels 2; Gaps 1;

```
Qy    3 AAATATCTGATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGAGAGGGCTTTGATGGC 62
Db  228 AAATGTGAAGTTTGCAGTTGAATATTTGTGCCTGGAACTTAGGAAAAACAATGTGTGGGC 169
Qy   63 GA--TACAGACTTGGAAGACATCAGTCGTGAGCAGTAAATGAGATGATTCAGGAAAGAGT 120
```

Db   168 CATCTGCATATCGGTGGTGATGAATGAAGCCACAGAAGCAGAAGCCAATCAATGCAGCAAGAGT 109
           |||||| ||||| |||| ||||||| || | ||| | |||||| ||||||

Qy   121 ATAAAACTGGGAAGAGAGACAGAGGACCAGCCTCAAGGAACACATATTTAAGGACTGGGTAGAAA 180

Db   108 AAATAAAGTGAAGACAAGAGAGTGAGGAGATTATTAGCA 10
           ||| ||| ||| ||||

Qy   181 AACAAGAGAGTATGAACAAGAGAGTGAGGAGATTATTAGCA 219

Db   48 AATGATAGTGTATTCAAATTGGCAAAAGAGACTGAGAA 10

RESULT 9
ABS41903/c
ID   ABS41903 standard; DNA; 300 BP.
XX
AC   ABS41903;
XX
DT   25-FEB-2003 (first entry)
XX
DE   Human liver single exon probe, SEQ ID No 16893.
XX
KW   Human; single exon nucleic acid probe; liver; cirrhosis;
KW   hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW   coronary heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157273-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000664.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488898/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human adult liver.
XX
PS   Claim 4; SEQ ID NO 16893; 658pp; English.
XX
CC   The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC   measuring human gene expression in a sample derived from human adult
CC   liver, comprising one of 13109 defined nucleotide sequences given in the
CC   specification (or complements/ fragments). The probe hybridises at high
CC   stringency to a nucleic acid molecule expressed in the human adult liver.
CC   (I) may be used for predicting, measuring and displaying gene expression
CC   in samples derived from human adult liver. The genes identified may be
CC   involved in genetic liver diseases such as cirrhosis,
CC   hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC   associated with coronary heart disease. ABS25011-ABS51005 represent human
CC   liver single exon nucleic acid probes of the invention. Note: The
CC   sequence information for this patent does not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 300 BP; 68 A; 87 C; 37 G; 108 T; 0 U; 0 Other;

Query Match   6.0%; Score 41.8; DB 4; Length 300;
Best Local Similarity 52.5%; Pred. No. 0.19;
Matches 115; Conservative 0; Mismatches 102; Indels 2; Gaps 1;

Qy   3 AAATATCTGATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGGAGAGGCTTTGATGGC 62
           |||||| ||||| |||||||| |||||||| | ||||| ||||||||||

Db   228 AAATGTGAAGTTTGCAGTTGAATATTTGTGCCTGGAACTTAGGAAAAACAATGTGTGGGC 169

Qy   63 GA--TACAGACTTGAAGACATCAGTGCTGAGCAGTAAATGAGATGATTCAGGAAAGAGT 120
           || || |||||||||||| ||||| |||| ||| ||||||||||||

Db   168 CATCTGCATATCGGTGATGAATGAAGCCACAGAAGCAGAAGCAATCAATGCAGCAAGAGT 109

Qy   121 ATAAACTGGGAAGAGAGACAGAGGACAGAGGCTCAAGGAACAACATATTTAAGGACTGGGTAGAAA 180
           ||||| || |||| ||| |||| |||| |||||||||

Db   108 AAATAAGTGAAGACAAGAGAGAAGAGGAAGAAAGAGAAGAAGAAGCCAATGTTTAAGGG 49

Qy   181 AACAAGAGAGTATGAACAAGAGAGTGAGGAGATTATTAGCA 219
           ||| ||| |||||| |||

Db   48 AATGATAGTGTATTCAAATTGGCAAAAGAGACTGAGAA 10

RESULT 10
ABS16349/c
ID   ABS16349 standard; DNA; 300 BP.
XX
AC   ABS16349;
XX
DT   19-AUG-2002 (first entry)
XX
DE   Human genome-derived single exon probe ORF from lung SEQ ID No 16340.
XX
KW   Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW   chronic obstructive pulmonary disease; interstitial lung disease;
KW   familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW   tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW   Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW   pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW   pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW   primary ciliary dyskinesis; pulmonary hypertension;
KW   hyaline membrane disease; open reading frame; ORF.
XX
OS   Homo sapiens.
XX
PN   WO200186003-A2.
XX
PD   15-NOV-2001.
XX
PF   30-JAN-2001; 2001WO-US000665.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2002-114183/15.
XX
PT   Spatially-addressable set of single exon nucleic acid probes, used to
PT   measure gene expression in human lung samples.
XX
PS   Claim 4; SEQ ID NO 16340; 634pp; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring single exon gene expression in a sample derived
CC   from human lung comprising single exon nucleic acid probes having one of
CC   12614 nucleic acid sequences mentioned in the specification, or their
CC   complements or the 12387 open reading frames derived from the 12614
CC   probes. Also included are a microarray comprising the novel set of probes
CC   ; the novel set of probes which hybridise at high stringency to a nucleic
CC   acid expressed in the human lung; measuring gene expression in a sample

CC   derived from human lung, comprising (a) contacting the array with a
CC   collection of detectably labeled nucleic acids derived from human lung
CC   mRNA, and (b) measuring the label detectably bound to each probe of the
CC   array; identifying exons in a eukaryotic genome, comprising (a)
CC   algorithmically predicting at least one exon from genomic sequences of
CC   the eukaryote; and (b) detecting specific hybridisation of detectably
CC   labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC   having a fragment identical to the predicted exon, the probe is included
CC   in the above mentioned microarray; assigning exons to a single gene,
CC   comprising (a) identifying exons from genomic sequence by the method
CC   above and (b) measuring the expression of each of the exons in several
CC   tissues and/or cell types using hybridisation to a single exon
CC   microarrays having a probe with the exon, where a common pattern of
CC   expression of the exons in the tissues and/or cell types indicates that
CC   the exons should be assigned to a single gene; a peptide comprising one
CC   of 12011 sequences, mentioned in the specification, or encoded by the
CC   probes/open reading frames (ORF). The probes are used for gene expression
CC   analysis, and for identifying exons in a gene, particularly using human
CC   lung derived mRNA and for the study of lung diseases such as asthma, lung
CC   cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC   disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC   tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC   Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC   histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC   Kartagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC   dyskinesis, pulmonary hypertension and hyaline membrane disease. The
CC   present sequence is a single exon probe open reading frame of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 300 BP; 68 A; 87 C; 37 G; 108 T; 0 U; 0 Other;

Query Match   6.0%; Score 41.8; DB 6; 37 G; 108 T; 0 U; 0 Other;
Best Local Similarity   52.5%; Pred. No. 0.19;
Matches   115; Conservative   0; Mismatches 102; Indels   2; Gaps   1;

Qy   3 AAAATCTGATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGAGAGGCTTTGATGGC 62
           |||| ||| ||| || || || |||| ||||||| || ||| || | || |||||
Db   228 AAATGTGAAGTTTGCAGTTGAATATTTGTGCCTGGAACTTAGGAAAAACAATGTGTGGC 169

Qy   63 GA--TACAGACTTGAAGACATCAGTGCTGCTGAGCAGTAAATGAGATGATTCAGGAAAGAGT 120
           || ||| ||| || || || || ||| | ||||| |||| || ||| ||||| ||
Db   168 CATCTCCATATCGGTGATGAATGAGAGCCACAGAAGCAGAGAGCAATCAATCAGCCAAGAGT 109

Qy   121 ATAAACTGGGAAGAGACAGAGACGCAGGCTTCAAGGAACATATTTAAGGACTGGGTAGAAA 180
           ||||| || || ||||| || | |||||||| || |||| ||| |||||| |||
Db   108 AAATAAAGTGAAGAAGAGGGAAGAGGAAGAGAGAAGAAGAGAGGAGCCAATGTTTAAGGG 49

Qy   181 AACAAGAGAGTATGAACAAGAGTGAGGAGATTATTAGCA 219
           || |||||| ||||| ||||| |||| |||| ||||| ||
Db   48 AATGATAGTGTATTCAAAATTGGCAAAAGAGACTGAGAA 10

RESULT 11
ABN64891
ID   ABN64891 standard; cDNA; 451 BP.
XX
AC   ABN64891;
XX
DT   28-JUN-2002 (first entry)
XX
DE   Human cancer related polynucleotide SEQ ID NO 4858.
XX
KW   Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW   gene therapy; cancer; tumour; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20021450-A2.
XX
PD   21-FEB-2002.
XX

PF   16-AUG-2001; 2001WO-US025840.
XX
PR   16-AUG-2000; 2000US-0226326P.
XX
PA   (CHIR) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
PI   Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
XX
DR   WPI; 2002-241905/29.
XX
PT   New nucleic acid for producing a polypeptide, detecting differentially
PT   expressed genes correlated with a cancerous state of a mammalian cell,
PT   and inhibiting tumor growth.
XX
PS   Claim 1; SEQ ID NO 4858; 883pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC   with cytostatic activity. The polynucleotide is used to produce a
CC   polypeptide, to detect differentially expressed genes correlated with a
CC   cancerous state of a mammalian cell and to inhibit tumour growth. The
CC   polynucleotide is used as a probe in mapping and tissue profiling. The
CC   encoded polypeptide and antibodies to the polypeptide can also be used
CC   for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC   gene therapy. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 451 BP; 129 A; 104 C; 106 G; 112 T; 0 U; 0 Other;

Query Match   5.9%; Score 41; DB 6; Length 451;
Best Local Similarity   61.9%; Pred. No. 0.37;
Matches   65; Conservative   0; Mismatches 40; Indels   0; Gaps   0;

Qy   10 TGATAGGCAGTTAGAAATTTGAGTTTGGAACACAGGAGAGAGGCTTTGATGGCGGATACAG 69
           ||| ||| ||| || || ||| || ||||| || || |||| || |||| || |||||
Db   72 TGTCAGGCAGATAGATCTATGAGTCTGGGAGCTCAGGCGGAGCTCTCGGGCTGAGATATAA 131

Qy   70 ACTTGGAAGACATCAGTGCTGCTGAGCAGTAAATGAGATGATTCAGGA 114
           ||||||| || || || | || || || || |||||| ||||| |
Db   132 ATTTGGAGTCATCAGCAGACAGCAGATGATATTTAAAGCCTGCAAGA 176

RESULT 12
AAQ96127/c
ID   AAQ96127 standard; DNA; 6305 BP.
XX
AC   AAQ96127;
XX
DT   07-MAR-1996 (first entry)
XX
DE   Tobacco leaf antifungal protein genomic DNA.
XX
KW   Tobacco; plant; fungicide; pesticide; antifungal protein; Phytophthora;
KW   transgenic plant; crop improvement; ds.
XX
OS   Nicotiana tabacum.
XX
FH   Key   Location/Qualifiers
FT   CDS   774..5743
FT   /*tag= r
FT   exon   774..911
FT   /*tag= a
FT   sig_peptide   774..839
FT   /*tag= p
FT   intron   912..1044
FT   /*tag= i
FT   exon   1045..1154
FT   /*tag= b
FT   intron   1155..3104
FT   /*tag= j
FT   exon   3105..3375

```
FT   intron           /*tag= c
FT                    3376..4225
FT                    /*tag= k
FT   exon             4226..4328
FT                    /*tag= d
FT   intron           4329..4431
FT                    /*tag= 1
FT   exon             4432..4777
FT                    /*tag= e
FT   intron           4778..4890
FT                    /*tag= m
FT   exon             4891..5253
FT                    /*tag= f
FT   intron           5254..5335
FT                    /*tag= n
FT   exon             5336..5414
FT                    /*tag= g
FT   intron           5415..5530
FT                    /*tag= o
FT   exon             5531..5743
FT                    /*tag= h
FT   polyA_signal     6095..6100
FT                    /*tag= q
FT                    /label= putative
XX
PN   WO9521929-A1.
XX
PD   17-AUG-1995.
XX
PF   95WO-EP000488.
XX
PR   94EP-00200321.
XX
PA   (MOGE-) MOGEN INT NV.
XX
PI   Melchers LS, Ponstein AS, Kroonswart S, Van Deventer- Troost JPE;
PI   Ohl SA, Bres-Vloemans AA, Logemann J, Sela-Buurlage MB;
XX
DR   WPI; 1995-293130/38.
DR   P-PSDB; AAR78679.
XX
PT   New antifungal protein and DNA obtd. from plants - used for producing
PT   transgenic plants or in antifungal compsns., partic. for retarding growth
PT   of Phytophthora infestans.
XX
PS   Disclosure; Page 27; 58pp; English.
XX
CC   This DNA sequence may be expressed a part of a chimeric DNA which is
CC   either expressed in a host cell or transgenic plant using binary vector
CC   plasmid pMOG846, or a replicon, preferably pMOG841. The resulting
CC   expressed protein displays fungicidal activity against Phytophthora
CC   infestans. The DNA may be expressed in a tissue- specific manner,
CC   especially in the style and stigma
XX
SQ   Sequence 6305 BP; 1905 A; 1094 C; 1114 G; 2192 T; 0 U; 0 Other;

Query Match             5.9%;  Score 40.8;  DB 2;  Length 6305;
Best Local Similarity  54.7%;  Pred. No. 1.1;
  Matches   81;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps   0;

Qy       72 TTGGAAGACATCAGTGCTGCTGAGCAGTAAATGAGATGAATATAAGAGAGTATAAACTGGGA 131
            ||||| | |||  |||| || |   |||| ||||  ||| |  || |||| ||   || ||
Db     3093 TTCCAAGATGTCAATTTAAATCTTTAAATAAGGTAATTCAGAAAATATTGTAATATAGTA 3034

Qy      132 AGAGGAGGAGGAGCAGGCTCAGGAGAACATATTTAAGGACTGGGTAGAAAAACAAGAGAGT 191
            |||   |   ||||  || || || |||| ||||||  ||||  |  ||   |||| |||
Db     3033 GAAGGGGCATTGATTTATCTAAATATTTAATATTTAATAATAAGAATTGTAGGGTCAGAAAAATAAT 2974

Qy      192 ATGAACAAGAGTGAGGAGATTATTAGCA 219
            ||  |||  |||  ||| ||| || ||
Db     2973 AGAAACAAAAGAGCTGAAATGCCAACA 2946
```

```
RESULT 13
ABL33696/c
ID   ABL33696/c standard; DNA, 6668 BP.
XX
AC   ABL33696;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1669.
XX
KW   Human; immune system disease; Cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1669; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6668 BP; 1628 A; 328 C; 1974 G; 2737 T; 0 U; 1 Other;

Query Match             5.8%;  Score 40.6;  DB 6;  Length 6668;
Best Local Similarity  50.3%;  Pred. No. 1.3;
  Matches  100;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps   0;

Qy      428 ACAACAGAAATAACAGATTTCCCAGCCTCTAACCAAAGAACAACAAGTTTGGGAACATT 487
            |||| ||||||| || ||||| ||| |||| ||||||| ||| ||| | || ||||||||
Db     5488 ACGAAAAAAAAAAAAAAAAACAAAACTTTCTAAAAAAAAAAAAACTCCAAACCAAT 5429

Qy      488 CCTCTCTTCTGAAATATGAAAGAGAGGGGATAAATACTGGAGTTAGGATTGTGAAAAAGT 547
            ||||||   |||| ||| || | ||| ||| | || ||  || ||   | ||| ||||
Db     5428 CTTACCAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5369

Qy      548 CAAGAGAAAAAAAAAGAACAGCCCAAGTGTAACAGATACTTCTTCCATGGGATGGTAAAA 607
            |||| ||||||||||||||
Db     5368 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACGAACGAAAAAAAAAAAAA 5309

Qy      608 GGAAGTTATTCAATAAAA 626
            ||||  | |||||||||
Db     5308 CAAACTTTCTAAAAAAAA 5290
```

```
RESULT 14
ACA28673
ID   ACA28673 standard; DNA; 1269 BP.
XX
AC   ACA28673;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #10330.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Clostridium botulinum.
XX
PN   WO20277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2002US-03429223P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,    Zamudio C,   Malone C,   Haselbeck R,   Ohlsen KL,   Zyskind JW;
PI   Wall D,    Trawick JD,  Carr GJ,    Yamamoto R,    Forsyth RA,   Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU24803.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 16543; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation;(7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC   prokaryotic essential genes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1269 BP; 498 A; 116 C; 217 G; 438 T; 0 U; 0 Other;

     Query Match             5.8%;  Score 40.2;  DB 7;  Length 1269;
     Best Local Similarity   47.6%;  Pred. No. 0.91;
     Matches  151;  Conservative   0;  Mismatches 163;  Indels    3;  Gaps    1;

Qy         19 GTTAGAAATTGAGTTTGGAACACCAGGAGAGAGGCTTTGATGGCGATACAGACTTGGAAG 78
              |||||||  |||||||||||| ||||  ||||||||| ||||||||   ||||||||||
Db        400 GCTTAAATATAGGCTATAAACATTCAAGTGGAGACTTTATAGCCGTATATGATGCAGAT 459

Qy         79 ACATCAGTGCTGAGCAGTAAATGAGATGATTCAGGAAAGAGTATAAACTGGGAAGAGGAC 138
              |||||||||||||  ||| || ||||||||| |||||||||||||||||||||||||||
Db        460 AATACTCCAGATAAAAATGCATTAAAATATCTTATGAAAACTATAATTGAGGACGAAGGT 519

Qy        139 AGAGGACAGGCTCAAGAACATATTTAAGGACTGGGTAGAAAAACAAGAGAGTATGAACA 198
              |||||||  |||||||||||||| ||||||||| |||| ||||| ||||||||||||||
Db        520 TTAGGAGCAGTTATAGGAA---AGTTTAGAACTAGAAATAAGACACAGAAATATGTTAACA 576

Qy        199 AGAGTGAGGAGATTATTAGCAGTGACCTTTGAGAGTACATCTCTAGAGTGTGGTATGTGTAG 258
              ||||||||  |||||| |||||||||||||  |||||||||||||||||||||||||||
Db        577 AGATTTATAAAATATAGAGACCTTAAGTTTTCAATGATGTGCCAAGCAGGAAGATGGAAG 636

Qy        259 AGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGGTGGAAGGAAAAGGAGCCAT 318
              ||||||||||||||||||||||||  ||||||||| ||||| |||| ||||||||||||
Db        637 TTATTAAATTTATGTACTATACCTGGCCACTAATTTTGTAGTAAGAAGAATATAATACAA 696

Qy        319 GAATCAAAAGGTCGGGA 335
              ||||||||||||||||
Db        697 GAGTTGAATGGATGGGA 713

RESULT 15
ABA17155/c
ID   ABA17155 standard; DNA; 32249 BP.
XX
AC   ABA17155;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 9486.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001334.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
```

```
PR   14-JUL-2000;   2000US-0218290P.
PR   26-JUL-2000;   2000US-0209963P.
PR   26-JUL-2000;   2000US-0209964P.
PR   14-AUG-2000;   2000US-0244518P.
PR   14-AUG-2000;   2000US-0244519P.
PR   14-AUG-2000;   2000US-0225213P.
PR   14-AUG-2000;   2000US-0225214P.
PR   14-AUG-2000;   2000US-0225266P.
PR   14-AUG-2000;   2000US-0225267P.
PR   14-AUG-2000;   2000US-0225268P.
PR   14-AUG-2000;   2000US-0225270P.
PR   14-AUG-2000;   2000US-0225447P.
PR   14-AUG-2000;   2000US-0225757P.
PR   14-AUG-2000;   2000US-0225758P.
PR   14-AUG-2000;   2000US-0225759P.
PR   18-AUG-2000;   2000US-0226279P.
PR   22-AUG-2000;   2000US-0226681P.
PR   22-AUG-2000;   2000US-0226868P.
PR   23-AUG-2000;   2000US-0271182P.
PR   30-AUG-2000;   2000US-0227009P.
PR   01-SEP-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   05-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   06-SEP-2000;   2000US-0230437P.
PR   06-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0230063P.
PR   14-SEP-2000;   2000US-0230064P.
PR   14-SEP-2000;   2000US-0230065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236227P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   20-OCT-2000;   2000US-0242221P.
PR   20-OCT-2000;   2000US-0244617P.
PR   01-NOV-2000;

PR   08-NOV-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251160P.
PR   05-DEC-2000;   2000US-0251198P.
PR   06-DEC-2000;   2000US-0256719P.
PR   08-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   11-DEC-2000;   2000US-0251990P.
PR   05-JAN-2001;   2000US-0254097P.
PR                  2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system cancers
PT   and metastases.
XX
PS   Disclosure; SEQ ID NO 9486; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA215534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
```

CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
XX   from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ   Sequence 32249 BP; 7365 A; 9042 C; 8692 G; 7150 T; 0 U; 0 Other;

Query Match   5.7%; Score 39.8; DB 5; Length 32249;
Best Local Similarity 46.9%; Pred. No. 3.9;
Matches 158; Conservative 0; Mismatches 177; Indels 2; Gaps 1;

Qy   43 AGGAGAGAGGCTTTGATGGCGATACAGACTTGGAGACATCAGTGCTGAGCAGTAAATGA 102
            ||||||||| || || || ||||| | ||| | |||| |
Db   20639 AGGAGAAAGAAGCAGGAGGAAGAGAAGAGGAGAAGAGAGAGAAGAGAGGGAGGAAAA 20580

Qy   103 GATGATTCAGGAAAGAGTATAAACTGGGAAGACAGGCTCAAGGAACATAT 162
               || | ||| | || ||| |||| || |||| |||| |
Db   20579 AGAGGAAGAAGAAGAGAAAAGAAGAGGAGAAGAGAGAAGCAGGAGGAGGAGGA 20520

Qy   163 TTAAGGACTGGGTAGAAAAACAAGAGAGTATGAACAAGAGTGAGGAGATTATTAGCAGTG 222
               ||||| || || |||||| || ||| ||||| |||||||
Db   20519 AGAGGGAGGAGAAGAAGAAAAGAGCAGAGGAGAAGAGGAGAAGAGAAGAGAAGGAATG 20460

Qy   223 ACCTTTGAGAGTACATCTCTAGAGTGGTATGTGTAGGAGCCAGATTTTC--AGCATCAGAA 280
               || ||| || | |||||| || | ||| | |||||
Db   20459 AGAAGGAGGAGGAGGAGGAAGGAGGAGGAGGAGGAAGAAGAGAAGAGAAGAAGAAGAA 20400

Qy   281 CCATCAAGCATTTTGGGGTGAAGGAAAAGAGCCATGAATCAAAAGGTGGGGAAAAGG 340
             ||||||| ||||| ||||| |||| || | |||| | ||| || |||| ||
Db   20399 GGAAGAGGAGCACAAGAAGAAGGAGAAAGAAGAGAAGAAGGAGAAGGGGGAGAACTG 20340

Qy   341 CCTTTTGGGAGTGGCAGTGTGGGTAGGGAGTAGGGC 377
              ||||||||| || || |||||||||| |||| |||
Db   20339 GAGTCCCTCAAGGGGAGCTGGAGCTCAGCATCGAGGTC 20303

Search completed: April 25, 2004, 08:45:59
Job time : 188.93 secs